Wavelet–Konstruktion als Anwendung der algorithmischen reellen algebraischen Geometrie

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
Herr Dipl.-Math. Lutz Lehmann
geboren am 30.10.1972 in Merseburg

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Wolfgang Coy

Gutachter:

1. Prof. Dr. Bernd Bank
2. Prof. Dr. Joos Heinz
3. Prof. Andreas Griewank, PhD

Tag der Verteidigung: 8. Februar 2007

Abstract

As a result of the TERA–project on Turbo Evaluation and Rapid Algorithms a new type, highly efficient probabilistic algorithm for the solution of systems of polynomial equations was developed and implemented for the complex case. The geometry of polar varieties allows to extend this algorithm to a method for the characterization of the real solution set of systems of polynomial equations.

The aim of this work is to apply an implementation of this method for the determination of real solutions to a class of example problems. Special emphasis was placed on the fact that those example problems possess real-life, practical applications. This requirement is satisfied for the systems of polynomial equations that result from the design of fast wavelet transforms.

During the last three decades discrete wavelet transforms arose as an important tool in signal analysis and in data compression (e.g. of picture or audio signals). The wavelet transforms considered here shall possess the practical important properties of symmetry and orthogonality. The specification of such a wavelet transform depends on a finite number of real parameters. Those parameters have to obey certain polynomial equations. If the system of those equations has real solutions at all, the solution set can consist of a finite number of points or can be a variety of positive dimension.

In the literature published on this topic, only example problems with a finite solution set were presented. For the computation of those examples it was sufficient to solve quadratic equations in one or two variables. This is easily done with the help of the tools of common computer algebra systems. However, the number of examples of this kind is rather small. Although there is an infinite number of example problems of this class with a finite complex solution set, it seems that there is only a small finite number among them that have real solutions.

Examples with real solution sets of positive dimension have the advantage that one can search for optimal solutions for some given, desired property. To characterize the set of real solutions of a system of polynomial equations it is a first step to find at least one point in each connected component. Already this is an intrinsically hard problem. The geometry of polar varieties ensures that, after a generic coordinate change, this problem can be transformed into a system of polynomial equations with a finite number of complex solutions. The thus obtained polynomial systems preserve the intrinsic hardness of the original problem and are thus a challenge for every computer algebra system. It turns out that the algorithm of the TERA–project performes very well with this task and is able to solve a larger number of examples than the best known commercial polynomial solvers.

Keywords:
real algebraic geometry, polar varieties, discrete wavelet transform, refinable functions

Zusammenfassung

Im Rahmen des TERA-Projektes (Turbo Evaluation and Rapid Algorithms) wurde ein neuartiger, hochgradig effizienter probabilistischer Algorithmus zum Lösen polynomialer Gleichungssysteme entwickelt und für den komplexen Fall implementiert. Die Geometrie polarer Varietäten gestattet es, diesen Algorithmus zu einem Verfahren zur Charakterisierung der reellen Lösungsmengen polynomialer Gleichungssysteme zu erweitern.

Ziel dieser Arbeit ist es, eine Implementierung dieses Verfahrens zur Bestimmung reeller Lösungen auf eine Klasse von Beispielproblemen anzuwenden. Dabei wurde Wert darauf gelegt, dass diese Beispiele reale, praxisbezogene Anwendungen besitzen. Diese Anforderung ist für die sich aus dem Entwurf von schnellen Wavelet-Transformationen ergebenden polynomialen Gleichungssysteme erfüllt.

Während der letzten drei Jahrzehnte entwickelten sich Wavelet-Transformationen zu einem bedeutenden Werkzeug der Signalanalyse und Datenkompression (z.B. von Bild– und Tonsignalen). Die hier betrachteten Wavelet-Transformationen sollen die praktisch wichtigen Eigenschaften der Orthogonalität und Symmetrie besitzen. Die Konstruktion einer solchen Wavelet-Transformation hängt von endlich vielen reellen Parametern ab. Diese Parameter müssen gewisse polynomiale Gleichungen erfüllen. Hat das System dieser polynomialen Gleichungen reelle Lösungen, so können diese in endlicher Anzahl vorkommen oder eine Mannigfaltigkeit positiver Dimension bilden.

In der veröffentlichten Literatur zu diesem Thema wurden bisher ausschließlich Beispiele mit endlichen Lösungsmengen behandelt. Zur Berechnung dieser Beispiele war es dabei ausreichend, quadratische Gleichungen in einer oder zwei Variablen zu lösen. Dies ist mit Unterstützung gewöhnlicher Computer-Algebra-Systeme leicht möglich. Jedoch ist die Anzahl der so bestimmbaren Beispiele recht klein. Es gibt zwar unendlich viele Beispiele in der betrachteten Problemklasse mit nur endlich vielen komplexen Lösungen, jedoch besteht Grund zu der Vermutung, dass nur eine kleine endliche Anzahl dieser Beispiele reelle Lösungen aufweist.

Beispiele mit reellen Lösungsmengen positiver Dimension haben den Vorteil, dass man nach optimalen Lösungen zu einer vorgegebenen gewünschten Eigenschaft suchen kann. Zur Charakterisierung der reellen Lösungsmenge eines polynomialen Gleichungssystems ist es ein erster Schritt, in jeder reellen Zusammenhangskomponente mindestens einen Punkt aufzufinden. Schon dies ist ein intrinsisch schweres Problem. Die Geometrie polarer Varietäten sichert, dass nach einer generischen Koordinatentransformation diese Aufgabe in ein polynomiales Gleichungssystem mit nulldimensionaler Lösungsmenge überführt werden kann. Das so erhaltene Gleichungssystem enthält aber weiterhin die intrinsische Komplexität des Ausgangsproblems, ist also eine Herausforderung für jedes Computer-Algebra-System. Es stellt sich heraus, dass der Algorithmus des TERA-Projektes zur Lösung dieser Aufgabe bestens geeignet ist und daher eine größere Anzahl von Beispielproblemen lösen kann als die besten kommerziell erhältlichen Lösungsverfahren.

Schlagwörter:
reelle algebraische Geometrie, polare Varietäten, diskrete Wavelet-Transformation, verfeinerbare Funktionen

Danksagung

Ich danke meinem Doktorvater Prof. Bernd Bank für seine Unterstützung und seine Bereitschaft sich auf das Teilgebiet der Wavelet-Theorie einzulassen und mit Rat und anregenden Diskussionen zum Abfassen dieser Arbeit beigetragen zu haben. Zusammen mit Prof. Carsten Carstensen danke ich beiden, dass sie mir den Rücken in organisatorischer Hinsicht freigehalten haben, so dass ich mich ganz auf die theoretische Arbeit konzentrieren konnte.

Den Mitgliedern des TERA-Projektes, insb. Joos Heintz in Santander und Buenos Aires, Marc Giusti, Gregoire Lecerf und Eric Schoost in Paris, Luis M. Pardo in Santander sowie Guillermo Matera, Rosita Wachenschauzer und Ariel Waissbein in Buenos Aires danke ich für ihre wissenschaftliche Vorarbeit und ihr Interesse an meinen Ergebnissen. Durch ihre großzügige Unterstützung bot sich mir mehrmals die Gelegenheit zum Halten von Vorträgen und zur konstruktiven Diskussion meiner Ergebnisse.

Ich war sehr erfreut, während dieser Vortragsreisen mit Stephane Mallat an der Ecole Polytechnique in Paris sowie mit Carlos Cabrelli an der Universidad de Buenos Aires zusammenzutreffen und danke ihnen für Diskussionen zu Fragen der Wavelet-Theorie.

Ganz besonders danken möchte ich Gregoire Lecerf und Eric Schoost für jenes wunderbare Stück Software, das sich „Kronecker-Package“ nennt und welches die in dieser Arbeit vorgenommenen Berechnungen erst möglich gemacht hat.

Den Sekretärinnen des Fachbereichs Optimierung und Diskrete Mathematik, Frau Ramona Klaass-Thiele und Jutta Kerger, gilt mein besonderer Dank für ihre vielfältige Hilfe in bürokratischen Fragen, die Korrektur englischer Manuskripte und ihre stete Sorge für das leibliche Wohlergehen in Form von Kaffee. Mein Dank gilt auch der engeren Arbeitsgruppe, insbesondere Bernd Wiebelt, Hergen Harnisch und Sebastian Heinz für ihre Bereitschaft zuzuhören und für die konstruktive und freundliche Arbeitsatmosphäre.

Meinen Eltern

Ich danke Euch, dass ihr meine Begabung erkannt habt und mich mit allen Euch zur Verfügung stehenden Mitteln gefördert habt.

Meiner Freundin Katina danke ich für die finale Rechtschreibkontrolle und für ihre Geduld, als das Schreiben dieser Arbeit sich länger als erwartet hinzog und die Freizeitgestaltung doch sehr darunter litt.

Inhaltsverzeichnis

Kapitel 1

Einleitung

Im Zuge der zunehmenden Verfügbarkeit hochleistungsfähiger Computer im Verlaufe der 1990er Jahre stieg auch das Interesse an der umfassenden Lösbarkeit von Problemen, die als nichtlineare Gleichungssysteme formuliert werden können. Für die klassischen numerischen Verfahren wie das Newton–Verfahren ist es wichtig, dass diese Gleichungssysteme nicht allzu sehr von linearen Gleichungssystemen abweichen bzw. dass man schon aus der Problemstellung einen guten Startwert für die Nullstelleniteration kennt.

1.1 Zu Lösungsverfahren polynomialer Gleichungssysteme

Für polynomiale Gleichungssysteme ist dies nicht der Fall. Seit den 1960er Jahren kennt man jedoch computerimplementierbare Verfahren, welche zumindest der Theorie nach eine vollständige Lösung solcher Gleichungssysteme erlauben. In diesen wird das polynomiale Gleichungssystem auf eine andere, wesentlich komplexere Art als dies in klassischen Lösungsverfahren der Fall ist, auf ein lineares Gleichungssystem sehr hoher Dimension zurückgeführt.

Dieses Verfahren der Gröbner– bzw. Standardbasen baut auf theoretischen Ansätzen auf, die bis zum Anfang des 19. Jahrhunderts zurückreichen. 1882 wurden durch L. Kronecker die existierenden Techniken zur Lösung spezieller polynomialer Systeme zu einer allgemeinen Lösungstheorie zusammengefasst. Durch D. Hilbert wurde diese Theorie weiter systematisiert, allerdings ersetzte er Kroneckers konstruktive und damit langen Beweise durch kurze, nichtkonstruktive Beweise. Die Theorie der Gröbner– bzw. Standardbasen entstand aus der Notwendigkeit, Hilberts nichtkonstruktive Resultate in konstruktive Methoden umzusetzen.

Im Verlaufe der 1980er Jahre und Anfang der 1990er Jahre wurden, beim Versuch, die Komplexität der Berechnung von Gröbner–Basen besser abzuschätzen bzw. diese Berechnung zu verbessern, die konstruktiven Methoden Kroneckers wiederentdeckt. Es stellte sich dabei heraus, dass man diese älteren Methoden wesentlich einfacher in ein computerimplementierbares Verfahren zum Lösen polynomialer Gleichungssysteme mit kontrollierbarem Ressourcenbedarf und Laufzeit entwickeln konnte. Eine Implementierung dieses *Kronecker–Algorithmus* getauften Verfahrens wurde Ende der 1990er Jahre erstellt.

Der fundamentale Unterschied, der diesen Fortschritt möglich machte, ist, dass bei der Bestimmung einer Gröbner–Basis die gesamte algebraische Struktur des Gleichungssystems er-

Tabelle 1.1: Tabelle der Laufzeiten

A	n	p	BKK	δ	δ^*	ktime	kmem	gtime	gmem	α	len
3	3	2	12	12	6	$3.2s$	1600 kB	$0.5s + 0.4s$	1200 kB	0.50	20
3	4	3	32	12	8	$7.2s$	2100 kB	$4s + 0.6s$	1700 kB	0.64	22
3	5	3	80	54	22	$170s$	4900 kB	$9900s + 2600s$	61700 kB/75200 kB	0.57	24
4	3	3	3	4	2	$1.8s$	3400 kB	$0s + 0.01s$	1300 kB	–	26
4	4	3	32	28	10	$23s$	2200 kB	$48.6s + 50.5s$	6500 kB	0.57	28
4	5	4	80	28	10	$42s$	3400 kB	$210s + 82s$	6800 kB/8400 kB	0.57	30
4	6	4	240	136	24	$2980s$	23300 kB	$> 22h$	$>$ 470000 kB	1.14	32
4	7	5	672	136	26	$5370s$	38000 kB	$> 10h$	$>$ 300 MB	–	34
5	4	3	32	32	6	$25s$	3400 kB	$153s + 225s$	8000 kB/13000 kB	0.65	32
5	5	4	80	32	10	$48s$	4500 kB	$420s + 325s$	9000 kB/15500 kB	0.46	34
5	6	4	240	168	36	$11853s$	65107 kB	$> 10h$	$>$ 300 MB	1.32	36

ComputerAlgebraSystem Magma

- Gemeinsame Parameter:
  - Approximationsordnung A und Anzahl freier Variablen n,
  - Anzahl p der Gleichungen zur Orthogonalität
  - BKK a–priori–Schranke für die Anzahl der Lösungen,
  - δ gefundene komplexe Lösungen,
  - darunter δ^* reelle Lösungen,
  - bester Hölder–Index α unter den Lösungen.
- Komplexitität des Kronecker–Algorithmus:
  - Zeit & Speicherbedarf für die GeometricSolve–Prozedur des Kronecker–Algorithmus
- Gröbner–Basen und Transformation auf Shape–Lemma–Form
  - Zeit & Speicherbedarf für die Groebner–Prozedur des CAS Magma, „grevlex“–Monomordnung
  - Zeit & Speicherbedarf für die ChangeOrder–Prozedur zu „lex“–Monomordnung und TriangularDecomposition–Prozedur

halten wird. Dies ist wesentlich mehr, als zum Zwecke der Nullstellenbestimmung benötigt wird. Es konnte gezeigt werden, dass z.B. viele aus geometrischen Problemen stammende Gleichungssysteme eine solch komplexe algebraische Natur besitzen, die jedoch für die reine Bestimmung der Nullstellen keine Bedeutung hat.

Die in dieser Arbeit dargestellten Resultate zur Konstruktion von Wavelet–Systemen entstanden im Rahmen des TERA–Projektes. Die Arbeit an diesen Wavelet–Systemen wurde mit der Zielstellung begonnen, eine Klasse praktisch sinnvoller Testbeispiele zu finden, mit welcher sich beide Verfahren, das auf den Gröbner–Basen basierende und der Kronecker–Algorithmus, vergleichen lassen (s. [LW01]). Die Ergebnisse waren von Anfang an sehr eindeutig. für Probleme mit vielen Nullstellen ist die TERA–Kronecker–Methode schneller bzw. als einzige überhaupt in der Lage, ein Ergebnis zu liefern.

1.2 Polynomiale Gleichungssysteme in der Wavelet–Theorie

Diskrete *Wavelet–Transformationen* sind eine Klasse orthogonaler, und damit umkehrbarer Transformationen, die reell– oder komplexwertige Folgen in Tupel von Folgen abbilden. Man wen-

det eine Wavelet–Transformation an, um qualitative Eigenschaften von Folgen, die als Modell *diskreter Signale* aufgefasst werden, besser charakterisieren zu können. Aus dieser Charakterisierung können die Signale je nach Anwendung näherungsweise oder exakt rekonstruiert werden. Für praktische Anwendungen benutzt man Wavelet–Transformationen, welche endliche reelle Folgen in Tupel endlicher reeller Folgen überführen. In diesem Fall besteht eine Wavelet–Transformation im wesentlichen aus einer Anzahl von Faltungen mit endlichen Folgen, die *Filter* genannt werden.

Wavelet–Transformationen zeichnen sich durch eine rekursive Struktur aus, siehe Abschnitt 3.5.1 auf Seite 97. Diese baut auf einer Grundstruktur auf, die (orthogonale) *Filterbank* genannt wird. Diese Filterbank bildet ebenfalls endliche reelle Folgen in Tupel endlicher reeller Folgen ab. Die rekursive Struktur der Wavelet–Transformation entsteht, wenn aus dem Bild der Filterbank die erste Folge herausgenommen und durch das Bild der Filterbank, angewandt auf diese Folge, ersetzt wird.

Eine Filterbank wird durch ganzzahlige und reelle Parameter charakterisiert; z.B. ist die Anzahl der reellen einer der ganzzahligen Parameter. Die reellen Parameter einer orthogonalen Filterbank sind nicht vollkommen frei wählbar. Sind die ganzzahligen Parameter fixiert, so ist das Tupel der reellen Parameter ein Punkte der reellen Nullstellenmenge eines Systems multivariater Polynome. Diese Polynome garantieren die Orthogonalität der zu den Parametern gehörigen Filterbank, siehe Abschnitt 3.2.2 auf Seite 87.

Diese Orthogonalitätsbedingungen sind leicht zu erfüllen. Ähnlich wie eine orthogonale Matrix in Spiegelungen zerlegbar ist, die wiederum durch Einheitsvektoren gegeben sind, ist auch eine orthogonale Filterbank durch eine Anzahl von Einheitsvektoren in einem reellen Spaltenvektorraum vollständig bestimmt, siehe Abschnitt 3.6.1 auf Seite 114.

Eines der Ziele, welche man bei Anwendung einer Wavelet–Transformation verfolgt, ist die einfachere Unterscheidung zwischen sich „langsam“ und sich „schnell“ ändernden Folgen. Man erhält aus einer orthogonalen Filterbank einen orthogonalen Projektor auf dem Folgenraum, indem man eine gegebene Folge zuerst mit der Filterbank transformiert, dann alle Folgen des entstehenden Tupels bis auf die erste durch die Nullfolge ersetzt und zuletzt auf dieses Tupel die inverse Transformation anwendet. Für eine „gute“ Filterbank soll dieser Projektor, angewandt auf sich „langsam“ ändernde Folgen, gute Approximationen dieser Folgen erzeugen. Die Güte einer orthogonalen Filterbank läßt sich danach bemessen, wie klein der Approximationsfehler auf einer Testklasse von sich „langsam“ ändernden Folgen ist.

Folgen, deren Änderungsgeschwindigkeit genau definierbar ist, sind die Schwingungsfolgen $e_\alpha := \{e_\alpha(n)\}_{n\in\mathbb{Z}}$ mit $e_\alpha(n) := e^{i(2\pi\alpha)n}$ und $\alpha \in \left[-\frac{1}{2}, \frac{1}{2}\right]$. Eine solche Folge ändert sich „langsam“ bei $\alpha \approx 0$ und „schnell“ bei $\alpha \approx \pm\frac{1}{2}$. Der Übergang von endlichen Folgen zu diesen unendlichen Folgen ist möglich, da die Faltung mit den endlichen Filterfolgen wie auch die weiteren Operationen, die eine Filterbank ausmachen, auch für beliebige unendliche Folgen definiert sind.

Die Güte und die Orthogonalität der Filterbank können gleichzeitig untersucht werden, wenn

man Folgen betrachtet, die sich als Koeffizienten einer Fourier–Entwicklung (s. Abschnitt B) ergeben, im Falle endlicher Folgen als Koeffizienten trigonometrischer Polynome, siehe Abschnitt 3.4.1 auf Seite 90. Man betrachtet also Folgen $c = \{c_n\}_{n\in\mathbb{Z}}$, welche eine Darstellung $c_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{c}(\alpha) e_\alpha(n)\,\mathrm{d}\alpha$ besitzen, wobei $\hat{c} \in L^2([-\frac{1}{2},\frac{1}{2}])$. Z.B. könnte $\hat{c} : \left[-\frac{1}{2},\frac{1}{2}\right] \to \mathbb{C}$ eine stückweise stetige Funktion sein. Funktion und Folge sind durch die Beziehung $\sum_{n\in\mathbb{Z}} |c_n|^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} |\hat{c}(\alpha)|^2\,\mathrm{d}\alpha$ verknüpft. Für die angegebene Klasse von Funktionen ist die rechte und damit die linke Seite endlich. Eine sich „langsam“ ändernde Folge ist in diesem Zusammenhang durch einen Träger $\operatorname{supp}\hat{c} \subset [-\varepsilon,\varepsilon]$ mit einem „kleinen“ Radius $\varepsilon > 0$ charakterisiert, eine „schnell“ oszillierende Folge durch $\operatorname{supp}\hat{c} \subset [-1,-1+\varepsilon] \cup [1-\varepsilon,1]$.

Aus der Forderung nach einfach zu erkennender Trennung schnell und langsam oszillierender Folgen gewinnt man weitere, lineare Gleichungen in den Parametern der Filterbank, sowie ein nichtlineares Fehlerfunktional, welches zu minimieren ist. Diese Gleichungen ergeben mit der oben erwähnten Parametrisierung orthogonaler Filterbänke ein polynomiales Gleichungssystem hohen Grades. Verknüpft man nun die hier erhaltenen Gleichungen mit den Identitäten, welche die Orthogonalität sichern, so kann die Anzahl von Gleichungen und Parametern reduziert werden, siehe Abschnitt 3.5.3 auf Seite 100.

Ersetzt man das nichtpolynomiale Gütekriterium durch ein schwächeres, aber quadratisches Funktional auf der Menge der orthogonalen Filterbänke, so ergibt sich ein polynomiales Optimierungsproblem mit einer quadratischen Zielfunktion und ebenso quadratischen Gleichungsnebenbedingungen, siehe Abschnitt 3.5.5 auf Seite 106.

Der Lagrange–Formalismus dieses Optimierungsproblems läßt sich in die Theorie der polaren Varietäten einbetten, siehe Abschnitt 2.6 auf Seite 59. Mit den innerhalb dieser Theorie verfügbaren Algorithmen können die reellen kritischen Punkte, und damit auch globale Minimalpunkte des Optimierungsproblems gefunden werden.

Genauere Einblicke in die Natur der gefundenen orthogonalen Filterbänke gewinnt man durch die Analyse, unter welchen Bedingungen es zu der Filterbank Skalierungs– und Wavelet-Funktionen gibt, die jeweils eine Multiskalenanalyse (siehe Abschnitt 5.3 auf Seite 150) und darauf aufbauend eine Waveletbasis erzeugen (siehe Abschnitt 5.3.5 auf Seite 161).

Numerische Ergebnisse zur Konstruktion von Waveletfilterbänken sind in Abschnitt 6.4 ab Seite 188 angegeben.

Kapitel 2

Zur Lösung polynomialer Gleichungssysteme und Optimierungsaufgaben

Wir werden später das Problem der Konstruktion „guter“ Wavelettransformationen untersuchen. Aus dieser Untersuchung ergibt sich, dass die Parameter jeder Wavelettransformation einem System polynomialer Gleichungen genügen müssen. Durch dieses sind diese Parameter jedoch noch nicht eindeutig festgelegt. In einfachen Fällen ergibt sich eine endliche Anzahl von Lösungen, im Allgemeinen hat jedoch die Nullstellenmenge des polynomialen Gleichungssystems eine positive Dimension.

Durch das Erweitern des Problems um ein weiteres Polynom, dessen Werte auf der Nullstellenmenge zu minimieren sind, und die Analyse des aus dem Lagrange–Ansatz folgenden Gleichungssystems kann ein erweitertes polynomiales Gleichungssystem gewonnen werden, welches wieder eine endliche Anzahl von Lösungen aufweist. Die reellen Punkte dieser Lösungsmenge beschreiben den reellen Teil der Nullstellenmenge, insbesondere können die Minimalpunkte der zusätzlichen Funktion bestimmt werden.

Aus der Untersuchung der „Güte“ einer Wavelettransformation ergeben sich Kandidaten für die zu minimierende Funktion. Diese sind jedoch nicht polynomial, meist nur Lipschitz–stetig. Diese Kandidaten der zu optimierenden Funktion können durch eine polynomiale Funktion nach oben abgeschätzt werden. Minimieren wir diese, so können wir erwarten, eine nahezu optimale Lösung im Sinne der Ausgangsfunktion zu erhalten.

Nachfolgend skizzieren wir die geometrisch–algebraischen Methoden zum Lösen polynomialer Gleichungssysteme $f_1, \ldots, f_n \in \mathbb{Q}[X_1, \ldots, X_n]$ mit endlich vielen isolierten Nullstellen $x \in \mathbb{C}^n$, $f_1(x) = \cdots = f_n(x) = 0$:

- das Eliminationsverfahren von Leopold Kronecker,
- das von Bruno Buchberger entwickelte, auf der Bestimmung von Gröbner–Basen beruhende Verfahren mit einer von Fabrice Rouillier entwickelten Methode, die eine Gröbner–Basis zur Bestimmung der Lösungen des Gleichungssystems auswertet, und
- das im Rahmen des von Joos Heintz initiierten TERA–Projekts (Turbo Evaluation – Rapid Algorithms) entworfene, von Gregoire Lecerf implementierte und vom Autor dieser Arbeit für das reelle Lösen algebraischer Probleme modifizierte Verfahren. Dieses greift

Ideen Kroneckers zur Elimination auf, reduziert jedoch die Komplexität auf das minimal Notwendige.

Obwohl das letztgenannte Verfahren noch in der Einführungs– und Erprobungsphase ist, scheint es für Testbeispiele mit hoher Anzahl n der Unbestimmten und hohem Grad d der Polynome effizienter in der Bestimmung der Nullstellen des Systems zu sein als das zweitgenannte Verfahren. Dies wird auch durch die Rechenexperimente dieser Arbeit belegt. Diese Tendenz ist auch nach den Komplexitätsaussagen (obere und untere Schranken) der jeweiligen Theorie der Verfahren zu erwarten. Einerseits gibt es für die Bestimmung von Gröbner–Basen lediglich exponentielle obere Schranken für den benötigten Arbeitsspeicher, doppelt exponentielle, scharfe obere Schranken für die Größe des Ergebnisses und ebenfalls doppelt exponentielle Schranken für die Rechenzeit (vgl. [GG99]). Auf der anderen Seite ist bekannt [Leh99, GLS01], dass diese Schranken für den TERA–Algorithmus in den die Ausgangsgleichung und deren Nullstellenmenge beschreibenden Größen n, d, δ polynomial von niedrigem Grade sind. Dabei sind wie zuvor $n \in \mathbb{N}_{>0}$ die Anzahl der Variablen, $d \in \mathbb{N}_{>1}$ eine gemeinsame Schranke für die Grade der Polynome des Ausgangssystems und δ eine Schranke für den *geometrischen Grad* der durch die Polynome des Systems definierten algebraischen Varietäten. Da dieser geometrische Grad einer Bézoutschen Ungleichung (s. [Hei79, HS80b, HS81, Hei83, Ful84, Vog84]) genügt, ist er durch $\delta \leq d^n$ beschränkt. Im schlechtesten Fall ergibt sich somit eine einfach exponentielle Laufzeit proportional zu $d^{O(n)}$.

Die Anwendung der TERA–Methode verlangt jedoch, dass das polynomiale Gleichungssystem bestimmte „gute“ Eigenschaften hat. Bei der Transformation des Lagrange–Ansatzes in ein Gleichungssystem gibt es gewisse frei wählbare Parameter. Die Auswirkungen dieser und die Bestimmung „guter“ Parameter lassen sich im Rahmen der Theorie polarer Varietäten untersuchen (vgl. [BGHM01, BGHP05] und die dort zitierte Literatur).

2.1 Einige Grundlagen der algebraischen Geometrie

Der Ausgangspunkt der algebraischen Geometrie ist die Theorie der *multivariaten Polynome* und der durch Systeme von Polynomen definierten Nullstellengebilde, der *algebraischen Varietäten*.

Von einem linearen Gleichungssystem mit rationalen Koeffizienten weiß man, dass Gleichungen, Linearkombinationen, d.h. die Summen von rationalen Vielfachen der Ausgangsgleichungen, ebenfalls von den Lösungen des Systems erfüllt sein müssen. Mit Hilfe des *Gaußschen Algorithmus* ist es möglich, solche Linearkombinationen aufzufinden, die die Struktur der Lösungsmenge genau charakterisieren.

Insbesondere im Fall eines linearen Gleichungssystems mit der gleichen Anzahl von Variablen und Gleichungen, dessen Lösungsmenge aus genau einem Punkt besteht, gibt es nach der *Cramerschen Regel* für jede Variable eine Linearkombination, die nur noch diese Variable enthält. Aus diesen Gleichungen kann also der Lösungspunkt bestimmt werden.

Um dasselbe für polynomiale Gleichungssysteme durchführen zu können, müssen Linear-

kombinationen betrachtet werden, die auch polynomiale Koeffizienten beinhalten. Diese bilden das *Ideal des Gleichungssystems*. Selbst unter der Bedingung, dass es Linearkombinationen gibt, welche nur noch eine Variable enthalten, stellt sich die Situation wesentlich komplizierter dar als im linearen Fall. Z.B. sind diese Linearkombinationen selbst wieder Polynome.

Neben diesen, gewissermaßen trivialen, Erweiterungen des ursprünglichen polynomialen Gleichungssystems kann man auch nichttriviale Erweiterungen durch Polynome betrachten, die ebenfalls für alle Punkte der algebraischen Varietät des Ausgangssystems verschwinden. Diese bilden wieder ein Ideal, das *Verschwindungsideal* der Varietät. Nach dem *Hilbertschen Nullstellensatz* ist es das *Radikal* zum Ideal des ursprünglichen Gleichungssystems. Enthält das Radikal Polynome, die nur von einer Variablen abhängen, so besitzen diese nur einfache Nullstellen.

Als Beispiel diene der Fall, dass die Lösungsmenge aus den Punkten $(0, 0)$ und $(1, 1)$ besteht. Man findet zwar, dass die Koordinaten den Gleichungen $X^2 - X = 0$ und $Y^2 - Y = 0$ genügen. Aber man kann aus den Lösungen beider Gleichungen vier Punkte zusammenstellen, von denen also zwei zu viel sind. In einem Fall mit mehr Variablen und mehr Lösungspunkten ergeben sich leicht Situationen, in welchen die Suche nach „echten“ Lösungen der nach der Nadel im Heuhaufen gleicht.

Andererseits läßt sich die Lösungsmenge meist auch nicht einfach auflisten, da die Koordinaten der Lösungspunkte nicht rational sein müssen. Es ist also eine Form der Lösung zu definieren, die auch nichtrationale Punkte mit endlichem Aufwand anzugeben vermag, die andererseits dabei auch keine zusätzlichen Punkte kodiert. Danach braucht es noch ein Verfahren, diese Lösung auch aus den gegebenen Gleichungen zu bestimmen.

Die Antwort auf die erste Frage ist die *geometrische Lösung* [GH91]. Diese Struktur beruht auf der Eliminationstheorie von Leopold Kronecker [Kro82], wonach es möglich ist, jeden isolierten Punkt der algebraischen Varietät durch eine *algebraische Zahl*, d.h. eine Nullstelle eines Polynoms mit rationalen Koeffizienten, zu parametrisieren. Die Koeffizienten des Lösungspunktes sind durch univariate Polynome mit rationalen Koeffizienten gegeben, die in der besagten algebraischen Zahl ausgewertet werden.

Besteht die algebraische Varietät nicht nur aus isolierten Punkten, so kann man diese in Teile zerlegen, so dass es für jeden dieser Teile lineare Koordinatenfunktionen gibt. Die Einschränkung eines der Teile der Varietät auf ein fixiertes Koordinatentupel besteht dann wieder aus endlich vielen isolierten Punkten.

Kroneckers Eliminationstheorie wurde mehrfach an die sich entwickelnde mathematische Begriffsbildung angepasst, von König [Kön03], Macaulay [Mac16] und in einer bis heute verständlichen Notation von van der Waerden [vW31].

Eine Antwort auf die zweite Frage wird in einem einfachen Fall, der aber schon die Struktur des allgemeinen Falls enthält, am Ende dieses Teils der Einleitung angegeben. Für den allgemeinen Fall sei auf Kapitel 2 verwiesen.

2.1.1 Multivariate Polynomringe

Wir setzen im Folgenden immer voraus, dass ein *Ring* ein Einselement, d.h. ein neutrales Element bzgl. der Multiplikation, enthält. Solche Ringe sind z.B. die ganzen Zahlen $\mathbb{Z}$ und ebenfalls die Restklassenringe $\mathbb{Z}/m\mathbb{Z}$, $m \in \mathbb{N}_{>1}$. *Körper* sind, wenn man die Eigenschaften der Division „vergisst", ebenfalls Ringe. Beispiele hierfür sind die rationalen, reellen und komplexen Zahlen, und auch die endlichen Körper $\mathbb{Z}/p\mathbb{Z}$, wenn p eine Primzahl ist.

Ein *Polynom* in $n \in \mathbb{N}_{>0}$ Variablen $X_1, \ldots, X_n$ mit Koeffizienten in einem Ring $\mathcal{R}$ läßt sich als eine Funktion $f : \mathcal{R}^n \to \mathcal{R}$,

$$x = (x_1, \ldots, x_n) \in \mathcal{R}^n \mapsto f(x) := \sum_{\alpha \in \mathbb{N}^n} f_\alpha x_1^{\alpha_1} \cdot \cdots \cdot x_n^{\alpha_n} ,$$

auffassen, welche durch eine endliche Koeffizientenfolge $\{f_\alpha\}_{\alpha \in \mathbb{N}^n}$ mit Gliedern in $\mathcal{R}$ gegeben ist. Eine Folge ist endlich, wenn nur endlich viele ihrer Glieder von Null verschieden sind, d.h. hier, wenn die Indexmenge $\{\alpha : f_\alpha \neq 0\}$ endlich ist. Daher ist die obige Reihe eine endliche Summe. Ein Polynom in einer einzigen Variablen wird *univariat* genannt, ein Polynom in mehreren Variablen *multivariat*.

Das Produkt der Potenzen der Koordinaten wird oft in Multiindexschreibweise als $x^\alpha := x_1^{\alpha_1} \cdot \cdots \cdot x_n^{\alpha_n}$ geschrieben. Ein Polynom, in dessen Koeffizientenfolge nur ein Glied von Null verschieden ist, wird *Monom* genannt. Ein Polynom kann somit auch als endliche Summe von Monomen charakterisiert werden.

Elementare Monome sind die Koordinatenabbildungen. Diese werden mit den Variablen identifiziert, es gelte also $X_k(x_1, \ldots, x_n) := x_k$, $k = 1, \ldots, n$. Damit können auch Monome in Multiindexschreibweise notiert werden, es sei $f_\alpha X^\alpha := f_\alpha X_1^{\alpha_1} \cdot \cdots \cdot X_n^{\alpha_n}$ dasjenige Monom, welches zum Index α den einzigen nicht verschwindenden Koeffizienten f_α besitzt.

Die *Summe* $h = f + g$ zweier Polynome f und g ist, da ebenfalls eine endliche Summe von Monomen, wieder ein Polynom. Dessen Koeffizientenfolge ergibt sich als gliedweise Summe der Koeffizientenfolgen der Summanden.

Da sich die Monome zweier Polynome nur auf endlich viele Weisen als Paare zusammenfassen lassen, und das Produkt zweier Monome wieder ein Monom ist, so ist auch das *Produkt* $h = f \cdot g$ zweier Polynome wieder ein Polynom. Dessen Koeffizientenfolge ergibt sich als *Faltung* der Koeffizientenfolgen der Faktoren,

$$h = \{h_\gamma\}_{\gamma \in \mathbb{N}^n} \quad \text{mit} \quad h_\gamma := \sum_{\alpha,\beta \in \mathbb{N}^n : \alpha + \beta = \gamma} f_\alpha \, g_\beta \, .$$

Man kann sich leicht davon überzeugen, dass die Polynome mit dieser Addition und Multiplikation die Axiome eines Rings erfüllen. Formal kann der Ring der Polynome ausgehend von den Koeffizientenfolgen definiert werden.

Definition 2.1.1 *Seien $n \in \mathbb{N}$, $\mathcal{R}$ ein Ring und*

$$\ell_{\text{fin}}(\mathbb{N}^n, \mathcal{R}) := \{f = \{f_\alpha\}_{\alpha \in \mathbb{N}^n} : f_\alpha = 0 \quad \textit{für fast alle} \quad \alpha \in \mathbb{N}^n\}$$

die Menge von endlichen Folgen *mit Gliedern in $\mathcal{R}$ und n Indizes.*

Der Polynomring $\mathcal{R}[\underline{X}] := \mathcal{R}[X_1, \ldots, X_n]$ *mit* Grundring $\mathcal{R}$ *und n* Variablen $X_1, \ldots, X_n$ *ist definiert als der Folgenraum $\ell_{\text{fin}}(\mathbb{N}^n, \mathcal{R})$ versehen mit der gliedweisen Addition und Multiplikation durch Faltung.*

Sei $|\alpha| := \alpha_1 + \cdots + \alpha_n$ für $\alpha \in \mathbb{N}^n$, man nennt $\deg X^\alpha := |\alpha|$ den Grad des Monoms X^α. Der Grad eines Polynoms $f = \sum_{\alpha \in \mathbb{N}^n} f_\alpha X^\alpha$ ist als größter Grad eines darin auftretenden Monoms definiert, also

$$\deg f := \max \{|\alpha| : \alpha \in \mathbb{N}^n \quad \& \quad f_\alpha \neq 0\}.$$

Ein Polynom in n Variablen des Grades $d := \deg f$ ist aus höchstens $\binom{n+d}{d}$ verschiedenen Monomen zusammengesetzt.

In einem Polynomring $\mathcal{R}[\underline{X}]$ ist der Ring $\mathcal{R}$ in Form der Monome vom Grad Null homomorph enthalten.

Definition 2.1.2 *Jeder Ringhomomorphismus $E : \mathcal{R}[\underline{X}] \to \mathcal{R}$, der $\mathcal{R}$ invariant läßt, wird Auswertungsabbildung genannt.*

Es gilt für beliebige $f \in \mathcal{R}[\underline{X}]$

$$E(f) = \sum_{\alpha \in \mathbb{N}^n} E(f_\alpha) E(X^\alpha) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha E(X_1)^{\alpha_1} \ldots E(X_n)^{\alpha_n},$$

die Auswertungsabbildung ist also durch ihre Werte $(E(X_1), \ldots, E(X_n)) \in \mathcal{R}^n$ in den Variablen vollständig bestimmt.

Umgekehrt kann zu jedem Punkt $x = (x_1, \ldots, x_n) \in \mathcal{R}^n$ der Ringhomomorphismus $E_x : \mathcal{R}[\underline{X}] \to \mathcal{R}$ definiert werden, der jedem Polynom $f \in \mathcal{R}[\underline{X}]$ den Wert

$$E_x(f) := f(x) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha x_1^{\alpha_1} \ldots x_n^{\alpha_n}$$

der Polynomfunktion f zuordnet. Insbesondere werden Monome aX^0 auf $a \in \mathcal{R}$ abgebildet und es gilt $E_x(X_k) = x_k$, $k = 1, \ldots, n$.

Ist $\mathcal{R}$ ein Körper der Charakteristik 0, z.B. $\mathcal{R} = \mathbb{Q}$ oder $\mathbb{R}$ oder $\mathbb{C}$, oder hat allgemein der Ring $\mathcal{R}$ unendlich viele Elemente, so gibt es für jedes von Null verschiedene Polynom wenigstens einen Punkt, in welchem es nicht zu Null ausgewertet wird. Insbesondere ist in diesem Fall die Beziehung zwischen Polynom und der durch das Polynom definierten Abbildung eineindeutig.

2.1.2 Polynomiale Gleichungssysteme, Ideale & Varietäten

Seien $\Bbbk$ ein Körper und $f_1, \ldots, f_s \in \Bbbk[X_1, \ldots, X_n]$ Polynome in n Variablen. Wir betrachten das mit diesen gebildetete Gleichungssystem

$$f_1(x) = \cdots = f_s(x) = 0$$

und wollen Nullstellen von diesem System bestimmen, d.h. Punkte $x \in \mathbb{k}^n$ mit

$$f_1(x) = \cdots = f_s(x) = 0\,.$$

Nun ist es oft erforderlich, dass die Nullstellenmenge wiederum Rückschlüsse auf das Gleichungssystem zuläßt. Dazu reichen die Punkte in $\mathbb{k}^n$ jedoch meist nicht aus.

Definition 2.1.3 *Ein Körper $\mathbb{k}$ heißt* algebraisch abgeschlossen, *wenn jedes Polynom $p \in \mathbb{k}[X]$ mindestens eine Nullstelle $a \in \mathbb{k}$ aufweist, d.h. $p(a) = 0$.*

Ist ein Körper $\mathbb{k}$ Teil eines algebraisch abgeschlossenen Körpers $\mathbb{K}$, so bezeichnet man als algebraischen Abschluss *von $\mathbb{k}$ in $\mathbb{K}$ die Menge $\overline{\mathbb{k}}$, die alle Elemente aus $\mathbb{K}$ enthält, die Nullstellen von Polynomen in $\mathbb{k}[X]$ sind.*

Beispiel: Aus dem Fundamentalsatz der Algebra folgt, dass der Körper $\mathbb{C}$ der komplexen Zahlen algebraisch abgeschlossen und der algebraische Abschluss des Körpers $\mathbb{R}$ der reellen Zahlen ist. Jedoch ist der algebraische Abschluss des Körpers $\mathbb{Q}$ der rationalen Zahlen in $\mathbb{C}$ eine echte Teilmenge $\overline{\mathbb{Q}} \subset \mathbb{C}$, welche keine transzendenten Zahlen enthält.

Bemerkung: Es kann für jeden Körper die Existenz eines algebraischen Abschlusses nachgewiesen werden. Allerdings beinhaltet dieser Nachweis im Allgemeinen Fall die Anwendung des Zornschen Lemmas. Für das praktische Rechnen mit Elementen des algebraischen Abschluss muss man sich daher auf einen Teil einschränken, der endlich konstruierbar ist.

Analog zur Folgerung aus dem Fundamentalsatz der Algebra zerfällt in einem algebraisch abgeschlossenen Körper jedes Polynom in Linearfaktoren.

Ist $x \in \overline{\mathbb{k}}^n$ eine Nullstelle des Gleichungssystems $f_1(x) = \cdots = f_s(x) = 0$, so ist x auch Nullstelle für polynomiale Vielfache der Ausgangspolynome und Summen davon. Die Menge der so erhaltenen Polynome nennt man ein *Ideal* des Polynomrings. Es ist häufig nützlich, wenn in diesem Ideal Polynome enthalten sind, die nur noch von einer der Variablen abhängen.

Definition 2.1.4 *Sei $\mathcal{R}$ ein Ring. Ein Ideal $I \subset \mathcal{R}$ ist eine Teilmenge, die unter Addition und Multiplikation mit beliebigen Ringelementen abgeschlossen ist. D.h., sind $g_1, \ldots, g_s \in I$ und $q_1, \ldots, q_s \in \mathcal{R}$, so muss deren Linearkombination im Ideal I enthalten sein,*

$$q_1 \cdot g_1 + \cdots + q_s \cdot g_s \in I\,.$$

Beispielsweise bilden die geraden Zahlen, die durch Drei teilbaren Zahlen etc. Ideale im Ring der ganzen Zahlen. Ein Körper besitzt nur triviale Ideale, den Körper selbst und das aus dem Nullelement bestehende Ideal.

In einem polynomialen Gleichungssystem $f_1, \ldots, f_s \in \mathbb{k}[X_1, \ldots, X_n]$ verschwinden alle Polynome des Ideals $I := \langle f_1, \ldots, f_s \rangle$ auf jeder Nullstelle des Gleichungssystems.

Definition 2.1.5 *Seien $\mathbb{k}$ ein Körper, $\overline{\mathbb{k}}$ ein algebraischer Abschluss von $\mathbb{k}$ und $I \subset \mathbb{k}[X_1, \ldots, X_n]$ ein Polynomideal. Die* (affine) algebraische Varietät *$V(I)$ zu I in $\overline{\mathbb{k}}^n$ ist die Nullstellenmenge des*

Systems aller Polynome von I,

$$V(I) := \{x \in \overline{\Bbbk}^n | \forall h \in I : h(x) = 0\} .$$

Die Nullstellenmenge des Gleichungssystems ist also die algebraische Varietät $V(\langle f_1, \ldots, f_s \rangle)$. Umgekehrt kann man sich die Aufgabe stellen, zu einer beliebigen Menge $M \subset \overline{\Bbbk}^n$ die kleinste $\Bbbk$–definierte algebraische Varietät zu finden, die diese Menge enthält, d.h. das größte Ideal in $\Bbbk[X_1, \ldots, X_n]$, dessen Polynome auf M verschwinden.

Definition 2.1.6 *Seien $\Bbbk$ ein Körper und $\overline{\Bbbk}$ ein algebraischer Abschluss von $\Bbbk$. Ist $M \subset \overline{\Bbbk}^n$ eine beliebige Teilmenge, so werde mit $I(M)$ das* Verschwindungsideal *in $\Bbbk[\underline{X}]$ bezeichnet, dessen Polynome in jedem Punkt von M eine Nullstelle haben,*

$$I(M) = \{h \in \Bbbk[\underline{X}] | \forall x \in M : h(x) = 0\} .$$

Wir können somit Ideale angeben, ohne von einer endlichen Menge von Polynomen auszugehen. Jedoch erhalten wir in einem Polynomring über einem Körper dadurch keine neuen Ideale, denn in diesem Fall sind alle Ideale durch endlich viele Polynome erzeugbar.

Satz 2.1.7 (Hilbertscher Basissatz, s. z.B. [vW31, GG99]) *Seien $\Bbbk$ ein Körper, dann ist jedes Ideal von $\Bbbk[X_1, \ldots, X_n]$ endlich erzeugt. Für jedes beliebige Ideal $I \subset \Bbbk[X_1, \ldots, X_n]$ gibt es also eine endliche Anzahl s von Polynomen $f_1, \ldots, f_s \in \Bbbk[X_1, \ldots, X_n]$, die das Ideal erzeugen, $I = \langle f_1, \ldots, f_s \rangle$.*

Man kann nun die Verknüpfung $I(V(I))$ betrachten, d.h. zu einem Ideal I wird die algebraische Varietät $V := V(I)$ konstruiert, zu dieser dann wieder das Verschwindungsideal $I(V)$. Gewiss gilt, dass, wenn die Potenz eines Polynoms f auf V verschwindet, dann verschwindet f auch auf V. Gehört also eine Potenz von f dem Ideal I an, so ist f in $I(V(I))$ enthalten.

Definition 2.1.8 *Sei $\mathcal{R}$ ein Ring. Ein Ideal $I \subset \mathcal{R}$ heißt* radikal, *wenn für jedes $a \in \mathcal{R}$ aus $a^n \in I$ für ein $n \in \mathbb{N}$ schon $a \in I$ folgt. Das Radikal $\sqrt{I}$ eines Ideals $I \subset \mathcal{R}$ ist das kleinste radikale Ideal, welches I enthält, genauer $\sqrt{I} = \{a \in \mathcal{R} \mid \exists n \in N : a^n \in I\}$.*

Da die ersten Potenzen eines Elements des Ideals im Ideal enthalten sind, gilt $I \subset \sqrt{I}$. Sind $a, b \in \sqrt{I}$, d.h. mit genügend großen $m, n \in \mathbb{N}$ gilt $a^m, b^n \in I$, so überlegt man sich leicht nach dem binomischen Lehrsatz, dass $(qa + pb)^{m+n} \in I$ gilt, also $\sqrt{I}$ wie oben definiert tatsächlich ein Ideal ist.

Lemma 2.1.9 *Sei $I \subset \Bbbk[\underline{X}]$ ein Ideal und $\sqrt{I}$ sein Radikal. Dann gilt $V(I) = V(\sqrt{I})$.*

Beweis: Da $I \subset \sqrt{I}$, sind alle Punkte von $V(\sqrt{I})$ in $V(I)$ enthalten. Sei $x \in V(I)$ und $h \in \sqrt{I}$ beliebig. Dann gibt es nach Definition 2.1.8 des Radikals ein $D \in \mathbb{N}$ mit $h^D \in I$, somit gilt $h(x)^D = 0$. Dies ist jedoch nur möglich, wenn $h(x) = 0$ gilt. Da h beliebig ist, ist x in $V(\sqrt{I})$ enthalten, also gilt die Gleichheit $V(I) = V(\sqrt{I})$. □

Ist $I = \Bbbk[\underline{X}]$, d.h. das konstante Polynom 1 ist im Ideal I enthalten, so kann kein Punkt aus $\overline{\Bbbk}^n$ eine Nullstelle aller Polynome im Ideal sein, die algebraische Varietät $V(I)$ ist leer. Diese Aussage gilt auch in umgekehrter Richtung.

Satz 2.1.10 (Hilbertscher Nullstellensatz, vgl. [vW31]) *Seien $\mathbb{k}$ ein Körper, $\overline{\mathbb{k}}$ ein algebraischer Abschluss und I ein Ideal in $\mathbb{k}[\underline{X}] = \mathbb{k}[X_1, \ldots, X_n]$. Dann ist $V(I)$ genau dann leer, wenn $1 \in I$, d.h. $I = \mathbb{k}[\underline{X}]$.*

Eine Folgerung aus dem Hilbertschen Nullstellensatz ist, dass das Ideal $I(V(I))$ einer algebraischen Varietät gerade das Radikal $\sqrt{I}$ ist.

2.1.3 Irreduzible Komponenten algebraischer Varietäten

Man kann zu einer beliebigen Punktmenge das Verschwindungsideal und zu dieser wieder die algebraische Varietät konstruieren, den *algebraischen Abschluss* der Punktmenge.

Definition 2.1.11 *Seien $\mathbb{k}$ ein Körper und $\overline{\mathbb{k}}$ ein algebraischer Abschluss. Sei weiter $M \subset \overline{\mathbb{k}}^n$ eine beliebige Teilmenge. Dann wird $\overline{M} := V(I(M))$, die Varietät zum Verschwindungsideal von M,* algebraischer *(bzw.* Zariski–Abschluss *bzgl. der* $\mathbb{k}$–Zariski–Topologie*) genannt.*

Der Begriff „*algebraische Varietät*“ ist also synonym zu „*algebraisch abgeschlossene Teilmenge*“. Der algebraische Abschluss einer Teilmenge einer algebraischen Varietät ist wieder eine Teilmenge dieser Varietät, denn das Verschwindungsideal der Teilmenge enthält das Ideal der Varietät.

Man kann nun eine gegebene algebraische Varietät $V \subset \overline{\mathbb{k}}^n$ in disjunkte Teile $V = C_1 \cup C_2$ zerlegen und von jedem Teil den algebraischen Abschluss betrachten. Die algebraischen Abschlüsse müssen nicht mehr disjunkt sein. Ist keine der Teilmengen vollständig in der Schnittmenge $\overline{C_1} \cap \overline{C_2}$ enthalten, so handelt es sich um eine *echte* Zerlegung der Varietät.

Definition 2.1.12 *Seien $\mathbb{k}$ ein Körper und $\overline{\mathbb{k}}$ ein algebraischer Abschluss. Sei weiter $V \subset \overline{\mathbb{k}}^n$ eine algebraische Varietät. Gilt für jede Zerlegung $V = C_1 \cup C_2$ in algebraische Varietäten $C_1, C_2 \subset \overline{\mathbb{k}}^n$ entweder $C_1 = V$ oder $C_2 = V$, so wird V* irreduzibel *genannt.*

Eine irreduzible algebraische Varietät läßt also keine echten Zerlegungen zu. Zum Beispiel ist der algebraische Abschluss eines Punktes aus $\overline{\mathbb{k}}^n$ irreduzibel, denn eine der Teilmengen einer Zerlegung enthält den Punkt und damit den gesamten algebraischen Abschluss. Somit kann man in jeder algebraischen Varietät V irreduzible Teilmengen finden. Jedoch sind diese meist zu klein in dem Sinne, dass sie im algebraischen Abschluss ihres Komplements vollständig enthalten sind.

Definition 2.1.13 *Seien $\mathbb{k}$ ein Körper, $\overline{\mathbb{k}}$ ein algebraischer Abschluss, $V \subset \overline{\mathbb{k}}^n$ eine algebraische Varietät und $C \subset V$ eine Teilmenge, die eine irreduzible Varietät ist. Ist der algebraische Abschluss $\overline{V \setminus C}$ eine echte Teilmenge von V, so wird C* irreduzible Komponente *von V genannt.*

Das Ideal $I(V)$ einer irreduziblen Varietät kann kein Polynom enthalten, dessen Faktoren sämtlich nur auf einem Teil von V verschwinden, da dies eine echte Zerlegung von V erzeugen würde. Mindestens einer der Faktoren muss also wieder $I(V)$ angehören.

Definition 2.1.14 *Sei $\mathcal{R}$ ein Ring. Ein Ideal $\mathfrak{p} \subset \mathcal{R}$ wird* prim *oder* Primideal *genannt, wenn aus $ab \in \mathfrak{p}$ eines von $a \in \mathfrak{p}$ oder $b \in \mathfrak{p}$ folgt.*

Man nennt ein Polynom *reduzibel*, wenn es in Faktoren zerlegt werden kann, sonst *irreduzibel*. In einem primen Polynomideal $\mathfrak{p}$ hat also jedes Polynom einen irreduziblen Faktor, der in $\mathfrak{p}$ enthalten ist. Insbesondere ist jedes Primideal radikal.

Als Folgerung aus dem *Hilbertschen Basissatz* erhält man, dass jedes Primideal $\mathfrak{p} \subset \Bbbk[\underline{X}]$ eines Polynomrings über einem Körper $\Bbbk$ von endlich vielen irreduziblen Polynomen erzeugt wird. Als weitere Folgerung ergibt sich, dass jedes radikale Ideal der Durchschnitt endlich vieler Primideale ist, also auch jede algebraische Varietät die Vereinigung einer endlichen Anzahl irreduzibler Komponenten ist.

2.1.4 Noether–Position irreduzibler Varietäten

Zur Motivation eines allgemeineren Dimensionsbegriffs algebraischer Varietäten sei eine nichtleere Teilmenge $M \subset \mathbb{R}^n$ eines reellen Spaltenvektorraums betrachtet. Mit einer der Standardtopologien von $\mathbb{R}^n$ ist auch M ein topologischer Raum. Sind nun stetige Funktionen $h_1, \ldots, h_r : M \to \mathbb{R}$ gegeben, so nennt man das Tupel $(h_1, \ldots, h_r)$ ein Koordinatensystem in einem Punkt $x \in M$, wenn es eine offene Umgebung $U \subset M$ von x und eine offene Teilmenge $W \subset \mathbb{R}^r$ gibt, so dass die Abbildung $h := (h_1, \ldots, h_r) : U \to W$ homöomorph ist. Gibt es in jedem Punkt von M ein Koordinatensystem aus r Funktionen, so nennt man r die *Dimension* von M.

Seien $\Bbbk$ ein Körper, $\overline{\Bbbk}$ ein algebraischer Abschluss und $V \subset \overline{\Bbbk}^n$ eine irreduzible Varietät. Es ist eine Folgerung der Kroneckerschen Lösungsmethode (vgl. [vW31] bzw. Abschnitt 2.3) für polynomiale Gleichungssysteme, dass es dann eine Anzahl $r \leq n$ von linearen Polynomen

$$L_k = a_{k,1}X_1 + \cdots + a_{k,n}X_n \in \Bbbk[\underline{X}] = \Bbbk[X_1, \ldots, X_n], \qquad k = 1, \ldots, r$$

gibt, die ein „gutes“ Koordinatensystem fast überall auf V definieren[1]. Dieses hat eine besonders einfache Form, wenn die linearen Polynome gerade die ersten r Variablen sind.

Definition 2.1.15 (Noether–Normalform, vgl. [GH91] für die Namensgebung) *Seien $\Bbbk$ ein Körper und $\overline{\Bbbk}$ ein algebraischer Abschluss. Eine algebraische Varietät $V \subset \overline{\Bbbk}^n$ befindet sich in* Noether–Normalform *der Dimension $r \leq n$, wenn*

- *kein Polynom aus $\Bbbk[X_1, \ldots, X_r]$ auf ganz V verschwindet und*
- *es für jedes $k = r+1, \ldots, n$ ein Polynom $q_k \in \Bbbk[X_1, \ldots, X_r][X_k] \cap I(V)$ gibt, dessen führender Koeffizient in X_k das Einselement von $\Bbbk[X_1, \ldots, X_r]$ ist.*

Durch diese Eigenschaften werden die ersten r Variablen zu dem, was oben „gutes Koordinatensystem“ genannt wurde. Durch eine lineare Transformation der Varietät kann jedes Tupel linearer Polynome auf die ersten Variablen transformiert werden und dann auf obige Eigenschaften untersucht werden.

Für $k = \mathbb{Q}$ und $\overline{\Bbbk} = \overline{\mathbb{Q}} \subset \mathbb{C}$ bedeuten die Anforderungen an eine Varietät V, um in Noether–Normalform zu sein, dass sie fast überall lokal durch die ersten r Koordinaten parametri-

[1] s. auch [GH91, HMW01] und dortige Quellen für Komplexitätsabschätzungen im Rahmen moderner effektiver Lösungsverfahren

sierbar ist. D.h. außerhalb einer Hyperfläche in $\overline{\mathbb{k}}^r$, d.h. der Nullstellenmenge eines einzelnen Polynoms, ist V eine endliche Überlagerung von $\overline{\mathbb{k}}^r$. Die ausgenommenen Punkte entstehen durch die Schnittpunkte der Blätter der Überlagerung, jedoch nicht durch Polstellen der Blätter. Insbesondere gibt es zu jeder kompakten Menge $K \subset \mathbb{C}^r$ eine kompakte Menge $L \subset \mathbb{C}^n$, so dass der Teil der Varietät V im Zylinder über K in der Menge L enthalten ist (vgl. [Leh99]).

Für eine beliebige Varietät müssen die guten Koordinatenfunktionen nicht aus den ersten Koordinaten bestehen.

Definition 2.1.16 (lineare Noether–Normalisierung von Varietäten) *Seien $\mathbb{k}$ ein Körper, $\overline{\mathbb{k}}$ ein algebraischer Abschluss und $V \subset \overline{\mathbb{k}}^n$ eine algebraische Varietät. Ein Koordinatenwechsel, der durch eine invertierbare Matrix $A \in \mathbb{k}^{n\times n}$ gegeben ist, heißt* Noether–Normalisierung *von V, wenn sich die transformierte Varietät*

$$V^A := A^{-1}V := \{y \in \overline{\mathbb{k}}^n : Ay \in V\}$$

in Noether–Normalform befindet.

Die transformierten Koordinaten haben die Form $y = A^{-1}x$. Die ersten r transformierten Koordinaten sind also lineare Polynome in $X = (X_1, \ldots, X_n)$, deren Koeffizienten Einträge der Matrix A^{-1} sind. Für eine Varietät, die in eine Noether–Normalform der Dimension r transformiert werden kann, stellen diese linearen Polynome „gute“ Koordinatenfunktionen dar.

Satz 2.1.17 (s. [GH91] und dort angegebene Quellen) *Seien $\mathbb{k}$ ein Körper der Charakteristik 0 bzw. mit unendlich vielen Elementen, und $\overline{\mathbb{k}}$ ein algebraischer Abschluss von $\mathbb{k}$. Dann kann für jede irreduzible algebraische Varietät $V \subset \overline{\mathbb{k}}^n$ ein Koordinatenwechsel bestimmt werden, der eine Noether–Normalisierung für V ist.*

Jeder irreduziblen Varietät kann eine Dimension zugeordnet werden, nämlich die Dimension ihrer Noether–Normalisierung.

Definition 2.1.18 *Eine algebraische Varietät V heißt* äquidimensional, *wenn alle irreduziblen Komponenten von V dieselbe Dimension besitzen.*

Die Aussage des Satzes zur Noether–Normalisierung kann dahingehend erweitert werden, dass fast jede invertierbare Matrix eine Noether–Normalisierung definiert. Die zu vermeidende Menge von Matrizen ist in einer Hyperfläche im Raum der Matrizen enthalten, d.h. in einer durch ein einziges Polynom definierten algebraischen Varietät. Die Vereinigung endlich vieler Hyperflächen ist wieder eine solche. Daher kann eine simultane Noether–Normalisierung für alle irreduziblen Komponenten einer äquidimensionalen algebraischen Varietät gefunden werden.

Weiter kann dieses Argument auf die simultane Betrachtung aller irreduziblen Komponenten einer beliebig gegebenen algebraischen Varietät ausgedehnt werden. Fast jede invertierbare Matrix definiert eine Koordinatentransformation, unter welcher sich jede irreduzible Komponente der algebraischen Varietät in Noether–Normalform der jeweiligen Dimension befindet.

2.1.5 Geometrischer Grad einer Varietät

Seien $\Bbbk$ ein Körper der Charakteristik 0, $\overline{\Bbbk}$ ein algebraischer Abschluss und $V \subset \overline{\Bbbk}^n$ eine irreduzible algebraische Varietät. Es gibt somit mindestens eine Noether–Normalisierung. Diese kann auch durch lineare Polynome $L_1, \ldots, L_r \in \Bbbk[X_1, \ldots, X_n]$ auf V angegeben werden, die voneinander linear unabhängig sind. Für jeden vorgegebenen Punkt $b \in \Bbbk^r$ definieren diese Polynome eine affine Ebene $V(L_1(X) - b_1, \ldots, L_r(X) - b_r)$ der Dimension $n - r$. Der Schnitt dieser Ebene mit der Varietät V besteht nach Definition der Noether–Normalisierung aus endlich vielen Punkten.

Definition 2.1.19 (s. [Hei79], [HS80b], [HS81], [Hei83], [Ful84, Vog84]) *Sei $V \subset \overline{\Bbbk}^n$ eine irreduzible algebraische Varietät der Dimension r. Sei H die Menge der affinen Ebenen $E \subset \overline{\Bbbk}^n$ der Dimension $n - r$, welche einen aus endlich vielen Punkten bestehenden Schnitt $E \cap V$ mit V besitzen. Der* geometrische Grad $\deg V$ *von V ist das Maximum der Anzahl der Punkte in $V \cap E$ über alle Ebenen $E \in H$.*

Sei $V \subset \overline{\Bbbk}^n$ eine beliebige algebraische Varietät. Der geometrische Grad $\deg V$ *von V ist dann die Summe der geometrischen Grade der irreduziblen Komponenten.*

Der geometrische Grad einer durch ein einzelnes Polynom $f \in \Bbbk[X_1, \ldots, X_n]$ definierten Hyperfläche $V(f)$ ist gerade der Grad von f als multivariates Polynom. Der geometrische Grad der Vereinigung zweier algebraischer Varietäten ist trivialerweise durch die Summe der geometrischen Grade beider Varietäten beschränkt. Für den Durchschnitt zweier algebraischer Varietäten gilt die *Bézout–Ungleichung*.

Satz 2.1.20 (s. [Hei79], [HS80b]) *Seien $\Bbbk$ ein Körper der Charakteristik 0, $\overline{\Bbbk}$ ein algebraischer Abschluss und $V, W \subset \overline{\Bbbk}^n$ zwei algebraische Varietäten. Dann gilt die* Bézout–Ungleichung

$$\deg V \cap W \leq \deg V \ \deg W \,.$$

2.2 Elementare Methoden der Algebra

Seien $\Bbbk$ ein Körper, $\overline{\Bbbk}$ ein algebraischer Abschluss von $\Bbbk$ und $I \subset \Bbbk[\underline{X}] = \Bbbk[X_1, \ldots, X_n]$ ein Polynomideal. Man kann die algebraische Varietät $V(I) \subset \overline{\Bbbk}^n$ mittels der auf V definierten Funktionen untersuchen. Dabei interessieren vor allem diejenigen Funktionen, welche sich als Polynome aus $\Bbbk[\underline{X}]$ fortsetzen lassen. Zwei Polynome definieren dieselbe Funktion auf $V(I)$, wenn ihre Differenz in I oder allgemeiner im Radikal $\sqrt{I}$ liegt.

Die polynomialen Funktionen auf V sind also Restklassen einer Äquivalenzrelation. Die Restklassenstruktur bleibt unter den arithmetischen Operationen Addition und Multiplikation erhalten, die Menge der Restklassen bildet also einen Ring, den *Koordinatenring* $\Bbbk[V]$ von V. Unter den Restklassen befinden sich auch diejenigen der konstanten Funktionen. Diese sind alle voneinander verschieden, wenn die Varietät nichtleer ist. Diese Einbettung von $\Bbbk$ in den Koordinatenring ergibt die Struktur einer $\Bbbk$–*Algebra*.

Befindet sich V in Noether–Normalform der Dimension r, so ist der Koordinatenring auch eine $\mathcal{R}$–Algebra über dem Ring $\mathcal{R} := \Bbbk[X_1, \ldots, X_r]$. Es kann dann gezeigt werden, dass es

zu jeder Funktion f des Koordinatenrings ein normiertes univariates Polynom $q \in \mathcal{R}[T]$ mit Koeffizienten in $\mathcal{R}$ gibt, so dass $q(f) = 0$ gilt. Elemente einer $\mathcal{R}$–Algebra, die eine solche polynomiale Identität erfüllen, werden *algebraisch* über $\mathcal{R}$ genannt.

Ist ein Punkt $P \in \overline{\Bbbk}^n$ gegeben, so bilden diejenigen Punkte der algebraischen Varietät V, deren erste r Koordinaten mit den Koordinaten von P übereinstimmen, die *Faser* von V über P. Wertet man die Koeffizienten von q in P aus, so erhält man ein univariates Polynom q_P. Unter den Nullstellen von q_P befinden sich die Werte von f in den Punkten der Faser von V über P.

Man kann dieses Verfahren zu einer Methode zum simultanen Auswerten mehrerer Funktionen des Koordinatenrings in den Punkten einer Faser von V verallgemeinern. Sind unter den Funktionen des Tupels auch die Klassen der Variablen $X_1, \ldots, X_n$, so erhält man aus der Untersuchung des Koordinatenrings eine Beschreibung der Punkte der Varietät, zusammen mit beliebigen weiteren Funktionswerten. Diese Beschreibung wird *geometrische Lösung* bzw. *Parametrisierung der Varietät à la Kronecker* genannt.

2.2.1 Restklassen– und Koordinatenring

Das Rechnen mit Restklassen ganzer Zahlen bzgl. Division mit Rest zu einem Divisor $m \in \mathbb{Z}$ kann auf beliebige Ringe und deren Ideale verallgemeinert werden. Die Menge $m\mathbb{Z}$ der Vielfachen von m ist ein Ideal in $\mathbb{Z}$, die Differenz von zwei ganzen Zahlen mit gleichem Rest bei Division durch m ist in diesem Ideal enthalten.

Seien $\mathcal{R}$ ein Ring und $I \subset \mathcal{R}$ ein Ideal. Analog zu den Restklassen ganzer Zahlen können zu I eine Äquivalenzrelation auf $\mathcal{R}$ und ein Restklassenring, der I aus $\mathcal{R}$ „herausfaktorisiert“, definiert werden.

Definition 2.2.1 *Seien $\mathcal{R}$ ein Ring und $I \subset \mathcal{R}$ ein Ideal. Der* Restklassenring $\mathcal{R}/I$ *ist die Menge der Äquivalenzklassen bzgl. der Relation $\sim_I$ mit $a \sim_I b$ gdw. $a - b \in I$. Wir bezeichnen die Restklasse von $a \in \mathcal{R}$ mit $a + I$. Die Abbildung $\varphi : \mathcal{R} \to \mathcal{R}/I$, welche jedem $a \in \mathcal{R}$ seine Restklasse $a + I$ zuordnet, heißt* Restklassenhomomorphismus.

Das Ideal I ist der Kern des Restklassenhomomorphismus. Generell ist der Kern eines jeden Ringhomomorphismus, ja das Urbild eines jeden Ideals unter einem Ringhomomorphismus wieder ein Ideal.

Definition 2.2.2 *Seien $\Bbbk$ ein Körper, $\overline{\Bbbk}$ ein algebraischer Abschluss von $\Bbbk$, $M \subset \overline{\Bbbk}^n$ eine beliebige Menge und $I(M) \subset \Bbbk[\underline{X}] = \Bbbk[X_1, \ldots, X_n]$ das Verschwindungsideal von M.*

Der Restklassenring $\Bbbk[M] := \Bbbk[\underline{X}]/I(M)$ zum Verschwindungsideal $I(M) \subset \Bbbk[\underline{X}]$ wird Koordinatenring *von M genannt.*

Der Koordinatenring einer Menge M stimmt mit dem Koordinatenring des algebraischen Abschluss überein, da die Verschwindungsideale übereinstimmen.

Lemma 2.2.3 *Seien $\Bbbk$ ein Körper, $\overline{\Bbbk}$ ein algebraischer Abschluss von $\Bbbk$, $I \subset \Bbbk[\underline{X}] = \Bbbk[X_1, \ldots, X_n]$ ein Polynomideal und $\mathcal{A} := \Bbbk[\underline{X}]/I$ dessen Restklassenring.*

- *I ist genau dann radikal, wenn $\mathcal{A}$ keine nichttrivialen nilpotenten Elemente enthält.*
- *I ist genau dann prim, wenn $\mathcal{A}$ keine nichttrivialen Nullteiler enthält.*

Beweis: Ist $f \in \Bbbk[\underline{X}]$ ein Polynom und $a := f + I$ seine Restklasse, so gilt $a = 0$ genau dann, wenn $f \in I$ gilt. a ist nilpotent genau dann, wenn es ein $N \in \mathbb{N}$ mit $a^N = 0$ gibt. Dies ist jedoch äquivalent zu $f^N \in I$, woraus die erste Behauptung folgt.

Ist $g \in \Bbbk[\underline{X}]$ ein weiteres Polynom mit Restklasse $b = g + I$, so ist $ab = 0$ äquivalent zu $fg \in I$. Daraus folgt die zweite Behauptung. □

2.2.2 Algebren und algebraische Elemente

Ist $\Bbbk$ ein Körper, so sind sowohl im Polynomring $\Bbbk[\underline{X}] = \Bbbk[X_1, \ldots, X_n]$ als auch in mit diesem gebildeten Restklassenringen Kopien von $\Bbbk$ in Form der konstanten Polynome und ihrer Restklassen enthalten. Dies kann auf allgemeine Ringe und in diesen enthaltene Teilringe verallgemeinert werden.

Definition 2.2.4 *Sei $\mathcal{R}$ ein Ring. Ein Ring $\mathcal{A}$ wird* $\mathcal{R}$–Algebra *genannt, wenn es einen injektiven Ringhomomorphismus $\iota : \mathcal{R} \to \mathcal{A}$ gibt.*

Eine $\mathcal{R}$–Algebra $\mathcal{A}$ enthält also einen zu $\mathcal{R}$ isomorphen Teilring. Man betrachtet zur Verkürzung der Notation den Ring $\mathcal{R}$ als identisch zu seiner isomorphen Kopie, d.h. $\mathcal{R} \subset \mathcal{A}$. Ein Homomorphismus von $\mathcal{R}$–Algebren ist ein Ringhomomorphismus, welcher den Grundring $\mathcal{R}$ invariant läßt.

Zum Beispiel ist ein Polynomring $\mathcal{R}[\underline{X}] := \mathcal{R}[X_1, \ldots, X_n]$ in n Variablen über einem Koeffizientenring $\mathcal{R}$ auch eine $\mathcal{R}$–Algebra. Ebenso ist dieser Polynomring $\mathcal{R}[\underline{X}]$ eine $\mathcal{R}[X_1, \ldots, X_k]$–Algebra für jedes $k = 1, \ldots, n$.

Unter den reellen bzw. komplexen Zahlen kann man diejenigen hervorheben, die Nullstellen von Polynomen mit ganzzahligen Koeffizienten sind. Die arithmetischen Rechenoperationen mit solchen Nullstellen führen wieder auf Nullstellen von Polynomen. Analog dazu kann man auch in einer beliebigen $\mathcal{R}$–Algebra $\mathcal{A}$ diejenigen Elemente betrachten, die Nullstellen von Polynomen mit Koeffizienten in $\mathcal{R}$ sind.

Definition 2.2.5 *Seien $\mathcal{R}$ ein Ring und $\mathcal{A}$ eine $\mathcal{R}$–Algebra.*

- *Ein Element $a \in \mathcal{A}$ wird* algebraisch *über $\mathcal{R}$ genannt, wenn es ein univariates Polynom $p \in \mathcal{R}[X]$ mit $d := \deg p > 0$ gibt, so dass $p(a) = 0$ gilt. Hat p als führenden Koeffizienten $p_d = 1$, so heißt p* normiert *und jedes $a \in \mathcal{A}$ mit $p(a) = 0$* algebraisch ganz *über $\mathcal{R}$.*
- *Eine $\mathcal{R}$–Algebra $\mathcal{A}$ heißt* algebraisch, *wenn jedes Element aus $\mathcal{A}$ algebraisch über $\mathcal{R}$ ist, sie heißt algebraisch ganz, wenn dies wieder für jedes Element gilt.*

Im Sinne der einleitenden Motivation ist $\sqrt{2}$ algebraisch ganz, $\sqrt{2/3}$ ist jedoch nur algebraisch über $\mathbb{Z}$, aber wieder algebraisch ganz über $\mathbb{Q}$. Ein Beispiel einer algebraisch ganzen $\mathbb{Z}$–Algebra

ist $\mathbb{Z}[\sqrt{2},\sqrt{3}]$. So hat $b := \sqrt{2}+\sqrt{3}$ das Quadrat $b^2 = 5+2\sqrt{6}$, d.h. wir erhalten $0 = (b^2-5)^2-24 = b^4-10b^2+1$ als normiertes Polynom zu b.

Sei $\mathcal{R}$ ein Ring. Polynome aus $\mathcal{R}[\underline{X}] := \mathcal{R}[X_1,\ldots,X_n]$ können dann in n–Tupeln aus jeder $\mathcal{R}$–Algebra $\mathcal{A}$ ausgewertet werden. Dies gestattet es, weitere Eigenschaften von $\mathcal{R}$–Algebren zu definieren.

Definition 2.2.6 *Sei $\mathcal{R}$ ein Ring und $\mathcal{A}$ eine $\mathcal{R}$–Algebra.*

- *$\mathcal{A}$ heißt* endlich erzeugt, *wenn es Elemente $a_1,\ldots,a_n \in \mathcal{A}$ gibt, so dass ganz $\mathcal{A}$ das Bild der Auswertungsabbildung*

$$E_{(a_1,\ldots,a_n)} : \mathcal{R}[X_1,\ldots,X_n] \to \mathcal{A}$$

  ist. Dies wird auch als $\mathcal{A} = \mathcal{R}[a_1,\ldots,a_n]$ notiert.

- *Ein Tupel $b_1,\ldots,b_r \in \mathcal{A}$ heißt* algebraisch frei *bzw.* algebraisch unabhängig, *wenn die Auswertungsabbildung*

$$E_{(b_1,\ldots,b_r)} : \mathcal{R}[Z_1,\ldots,Z_r] \to \mathcal{A}$$

  injektiv ist, d.h. $\mathcal{A}$ mittels dieser Abbildung zu einer $\mathcal{R}[Z_1,\ldots,Z_r]$–Algebra wird.

Der Polynomring $\mathcal{R}[\underline{X}] := \mathcal{R}[X_1,\ldots,X_n]$ ist also selbst schon eine endlich erzeugte $\mathcal{R}$–Algebra, ebenso eine endlich erzeugte $\mathcal{R}[X_1]$–Algebra, etc. Die Variablen des Polynomrings sind trivialerweise algebraisch frei.

Ist $\Bbbk$ ein Körper und $I \subset \Bbbk[\underline{X}] = \Bbbk[X_1,\ldots,X_n]$ ein Polynomideal, so ist der Restklassenring $\Bbbk[\underline{X}]/_I$ eine endlich erzeugte $\Bbbk$–Algebra, erzeugt von den Restklassen der Variablen $X_1,\ldots,X_n$. Ist $\overline{\Bbbk}$ ein algebraischer Abschluss und $V \subset \overline{\Bbbk}^n$ eine algebraische Varietät in Noether–Normalform der Dimension r, so ist der Koordinatenring nach Definition eine $\Bbbk[X_1,\ldots,X_r]$–Algebra. Aus der zweiten Eigenschaft der Definition folgt, dass die Restklassen der Variablen $X_1,\ldots,X_n$ algebraisch ganz über $\Bbbk[X_1,\ldots,X_r]$ sind. Im folgenden wird ersichtlich, dass daraus schon folgt, dass der Koordinatenring $\Bbbk[V]$ algebraisch ganz über $\Bbbk[X_1,\ldots,X_r]$ ist.

Um zu dieser Schlussfolgerung zu gelangen und nachfolgend Lösungen eines polynomialen Gleichungssystems $f_1,\ldots,f_n \in \mathbb{Q}[X_1,\ldots,X_n]$ als Tupel algebraischer Zahlen aus $\mathbb{C}$ darzustellen, müssen die Rechenoperationen mit algebraischen Zahlen genauer spezifiziert werden. Meist kann man diese nicht direkt angeben, da jede Darstellung eine unendliche Länge hat. Jedoch ist jede algebraische Zahlen in $\mathbb{C}$ Nullstelle eines univariaten Polynoms. Man kann diese daher indirekt durch dieses Polynom und ein Rechteck in $\mathbb{C}$ mit rationalen Eckpunkten eindeutig charakterisieren.

Um z.B. die Summe zweier solcherart indirekt durch Polynome p,q gegebener algebraischer Zahlen a,b zu bestimmen, muss ein Polynom gefunden werden, welches die Summe beider als Nullstelle besitzt. Eine Möglichkeit ist, aus den Polynomen $p(X)$ und $q(Y-X)$ die gemeinsame Variable X zu eliminieren, denn mit $X = a$ und $Y = a+b$ verschwinden beide Polynome.

2.2.3 Faktorielle Integritätsbereiche

Mit univariaten Polynomen über einem Körper kann man ähnlich rechnen wie mit ganzen Zahlen. So kann man mittels des *Euklidischen Algorithmus* das größte gemeinsame Vielfache zweier univariater Polynome wie auch zweier ganzer Zahlen bestimmen, man kann univariate Polynome wie auch ganze Zahlen eindeutig in Primfaktoren zerlegen.

Man kann die Voraussetzungen für diese Konstruktionen als Anforderungen an Ringe formulieren, die sich auf univariate Polynomringe über den so spezifizierten Ringen vererben.

Eine erste wichtige Eigenschaft ist, dass die Gradarithmetik der univariaten Polynome sich „normal“ verhält, d.h. dass der Grad eines Produkts die Summe der Grade der Faktoren ist. Dazu ist zu fordern, dass kein Produkt von Ringelementen die Null ergibt, es sei denn, ein Faktor ist Null.

Definition 2.2.7 *Sei $\mathcal{R}$ ein Ring. Ein Ringelement $a \neq 0$ wird* Nullteiler *genannt, wenn es ein weiteres Ringelement $b \neq 0$ mit $ab = 0$ gibt.*

Ist $\mathcal{R}$ ein Ring mit Eins ohne Nullteiler, so wird $\mathcal{R}$ Integritätsbereich *genannt.*

Beispielsweise sind die Klassen von $a, b \in \mathbb{Z}$ in $\mathbb{Z}/(ab)\mathbb{Z}$ Nullteiler. Körper sind immer Integritätsbereiche.

Lemma 2.2.8 *Ist $\mathcal{R}$ ein Integritätsbereich, so auch der univariate Polynomring $\mathcal{R}[Y]$.*

Beweis: Seien $a, b \in \mathcal{R}[Y]$ beide von Null verschieden. Dann muss das Produkt $a_{\deg a} b_{\deg b}$ der führenden Koeffizienten von Null verschieden sein. Da dieses Produkt dann der führende Koeffizient des Produkts der Polynome ist, ist dieses ebenfalls von Null verschieden. Es gibt also keine Nullteiler in $\mathcal{R}[Y]$. □

Dies kann per Induktion auf mehrere Variable ausgedehnt werden, da der Polynomring $\mathcal{R}[X_1, \ldots, X_n]$ in n Variablen zum Polynomring $\mathcal{R}[X_1, \ldots, X_{n-1}][X_n]$ in einer Variablen mit Grundring $\mathcal{R}[X_1, \ldots, X_{n-1}]$ kanonisch isomorph ist.

Um nun mit den univariaten Polynomen über einem Integritätsbereich zu rechnen, als ob ihre Koeffizienten Elemente eines Körpers wären, kann man versuchen, den Integritätsbereich in einen Körper einzubetten. Ein passender Körper $Quot(\mathcal{R})$ läßt sich analog zur Konstruktion des Körpers der rationalen Zahlen $\mathbb{Q}$ aus dem Integritätsbereich $\mathbb{Z}$ der ganzen Zahlen konstruieren.

Definition 2.2.9 *Sei $\mathcal{R}$ ein Integritätsbereich. Der* Quotientenkörper *$\mathcal{K} := Quot(\mathcal{R})$ ist die Menge der Äquivalenzklassen der Relation $\sim$ über der Menge der Paare $\{(a, b) \in \mathcal{R}^2 : b \neq 0\}$, wobei*

$$(a_1, b_1) \sim (a_2, b_2) \iff a_1 b_2 = a_2 b_1 \, .$$

Die Klasse des Paars (a, b) wird als Bruch $\frac{a}{b}$ notiert.

$\mathcal{R}$ ist in $Quot(\mathcal{R})$ als Menge der Brüche $\frac{a}{1}$ enthalten. Ist $\mathcal{R}$ ein Integritätsbereich, so nach Lemma 2.2.8 auch jeder Polynomring $\mathcal{R}[\underline{X}] = \mathcal{R}[X_1, \ldots, X_n]$.

Definition 2.2.10 *Ist $\mathcal{R}$ ein Integritätsbereich, so wird der Quotientenkörper des Polynomrings $\mathcal{R}[\underline{X}] = \mathcal{R}[X_1, \ldots, X_n]$ als Körper der rationalen Funktionen bezeichnet und mit $\mathcal{R}(\underline{X}) := \mathcal{R}(X_1, \ldots, X_n) := Quot(\mathcal{R}[\underline{X}])$ notiert.*

Um vom Quotientenkörper $Quot(\mathcal{R})$ wieder zum Integritätsbereich zurückzugelangen, benötigt man die Existenz des größten gemeinsamen Teilers bzw. des kleinstem gemeinsamen Vielfachen eines Tupels von Ringelementen. Um diese zu garantieren, muss man die starke Forderung nach einer eindeutigen Primfaktorzerlegung aller Ringelemente stellen. Diese Eindeutigkeit der Primfaktorzerlegung ist relativ zum Austausch von invertierbaren Ringelementen zu betrachten.

Definition 2.2.11 (vgl. [GG99], Anhang) *Sei $\mathcal{R}$ ein Integritätsbereich. Ein Element $e \in \mathcal{R}$ wird* Einheit *genannt, wenn es ein Element $f \in \mathcal{R}$ mit $ef = 1$ gibt, d.h. wenn e im Ring invertierbar ist.*

Man nennt ein Element $a \in \mathcal{R}$ Teiler von $b \in \mathcal{R}$, notiert $a|b$, wenn es ein weiteres Ringelement $c \in \mathcal{R}$ mit $ac = b$ gibt.

Ein Element p des Rings heißt prim, *wenn für alle $a, b \in \mathcal{R}$ aus $p|ab$ eines von $p|a$ oder $p|b$ folgt. D.h. das von p erzeugte Ideal $\langle p \rangle$ ist ein Primideal von $\mathcal{R}$.*

$\mathcal{R}$ heißt faktorieller Integritätsbereich *oder* Ring mit eindeutiger Primfaktorzerlegung, *wenn es für jedes $a \in \mathcal{R}$ eine Anzahl $n \in \mathbb{N}$ von Primelementen $p_1, \ldots, p_n \in \mathcal{R}$ mit*

$$a = p_1 \cdots p_n$$

gibt und alle Zerlegungen von a in Primfaktoren bis auf Reihenfolge und Multiplikation der Faktoren mit Einheiten übereinstimmen.

Somit kann in einem faktoriellen Ring zu jeder endlichen Anzahl von Ringelementen ein – bis auf eine Einheit – eindeutiger größter gemeinsamer Teiler gefunden werden, indem man die Primfaktorzerlegungen der Elemente vergleicht. Ebenso ist es möglich, ein kleinstes gemeinsames Vielfaches zu erhalten.

Insbesondere kann man von einem Polynom mit Koeffizienten in einem faktoriellen Integritätsbereich den größten gemeinsamen Teiler, den *Inhalt* des Polynoms, abspalten. Das verbleibende Polynom wird *primitiver Teil* des Polynoms genannt. Ein Polynom, dessen Inhalt eine Einheit ist, wird *primitiv* genannt. Der Inhalt des Produkts zweier Polynome ist das Produkt der Inhalte der Faktoren. In einer Identität von Ausdrücken von Polynomen kann also der Inhalt beider Seiten getrennt von den primitiven Anteilen betrachtet werden.

Satz 2.2.12 (Satz von Gauß, vgl. [GG99], Satz 6.8) *Sei $\mathcal{R}$ ein faktorieller Integritätsbereich. Dann ist der univariate Polynomring $\mathcal{R}[X]$ ebenfalls ein faktorieller Integritätsbereich.*

Ist also $\mathcal{R}$ ein faktorieller Integritätsbereich, und sind $p_0, p_1 \in \mathcal{R}[X]$ univariate Polynome mit Koeffizienten in $\mathcal{R}$, so existiert der größte gemeinsame Teiler $q := ggT(p_0, p_1)$. Als Polynome

mit Koeffizienten im Quotientenkörper $\mathcal{K} := Quot(\mathcal{R})$ kann der euklidische Algorithmus ausgeführt werden. Dabei werden Polynome $q_1, \ldots, q_N \in \mathcal{K}[X]$ und $p_2, \ldots, p_{N+1} \in \mathcal{K}[X]$ mit

$$p_{k+1} := p_{k-1} - q_k p_k, \quad \deg p_{k+1} < \deg p_k$$

konstruiert, so dass $p_N \neq 0$ und $p_{N+1} = 0$ gelten. Fasst man die Polynome zu Spaltenvektoren zusammen, so gilt

$$\begin{pmatrix} p_k \\ p_{k+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q_k \end{pmatrix} \begin{pmatrix} p_{k-1} \\ p_k \end{pmatrix},$$

zwischen ersten und letzten Vektor also

$$\begin{pmatrix} p_N \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q_N \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & -q_1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}.$$

Es gibt somit Polynome $a_0, a_1 \in \mathcal{K}[X]$ mit $a_0 p_0 + a_1 p_1 = p_N$. Die auftretenden Matrizen sind invertierbar, für die Beziehung zwischen erstem und letzten Spaltenvektor gilt ebenfalls

$$\begin{pmatrix} p_0 \\ p_1 \end{pmatrix} = \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} q_N & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_N \\ 0 \end{pmatrix},$$

p_N teilt somit sowohl p_0 als auch p_1. Es folgt, dass p_N ein Vielfaches des größten gemeinsamen Teilers von p_0 und p_1 sein muss. Durch Bilden des Hauptnenners und Abgleichen der Inhalte beider Seiten von $a_0 p_0 + a_1 p_1 = p_N$ kann $q = ggT(p_0, p_1) \in \mathcal{R}[X]$ bestimmt werden, es gibt weiterhin Polynome $a_0, a_1 \in \mathcal{R}[X]$ mit $a_0 p_0 + a_1 p_1 = q$.

2.2.4 Minimalpolynome

Ist $\mathcal{R}$ ein Ring und $\mathcal{A}$ eine $\mathcal{R}$–Algebra, so kann für jedes $a \in \mathcal{A}$ die Auswertungsabbildung $\varphi : \mathcal{R}[X] \to \mathcal{A}$ mit $X \mapsto a$ definiert werden. Ist a algebraisch über $\mathcal{R}$, so ist der Kern dieser Abbildung ein nichttriviales Ideal $I \subset \mathcal{R}[X]$. Eine besonders einfache Situation entsteht, wenn das Ideal von einem der gradminimalen Elemente des Ideals erzeugt wird.

Definition 2.2.13 *Seien $\mathcal{R}$ ein Ring, $\mathcal{A}$ eine $\mathcal{R}$–Algebra und $a \in \mathcal{A}$ algebraisch über $\mathbb{k}$. Wird das Ideal $I := \{p \in \mathcal{R}[Y] : p(a) = 0\}$ von einem gradminimalen Polynom $q \in \mathcal{R}[Y]$ erzeugt, $I = \langle q \rangle$, so wird q* Minimalpolynom *von a genannt.*

Lemma 2.2.14 *Seien $\mathcal{R}$ ein faktorieller Integritätsbereich und $I \subset \mathcal{R}[X]$ ein Ideal univariater Polynome. Gibt es in I ein normiertes Polynom p, so wird I von einem gradminimalen normierten Polynom q erzeugt, $I = \langle q \rangle$.*

Beweis: Seien a ein gradminimales Polynom und $q := ggT(a, p)$ der größte gemeinsame Teiler von a und dem normierten Polynom p. q muss als Faktor von p ebenfalls normiert sein. Da a gradminimal ist, muss q denselben Grad haben, q unterscheidet sich nur um einen konstanten Faktor von a. Nach dem erweiterten Euklidischen Algorithmus kann q weiter als Linearkombination von a und p dargestellt werden, gehört also ebenfalls dem Ideal I an.

Jedes weitere $b \in I$ muss ebenfalls ein Vielfaches von q sein, da das Polynom $ggT(q, b) \in I$ sonst einen kleineren Grad als q besitzen müsste. □

Definition 2.2.15 *Wenn alle Ideale eines Rings von je einem Ringelement erzeugt werden, so nennt man diesen Ring einen* Hauptidealring.

$\mathcal{R}[X]$ ist also für jeden faktoriellen Integritätsbereich $\mathcal{R}$ ein Hauptidealring. Ist $\mathcal{A}$ eine $\mathcal{R}$–Algebra und ist $a \in \mathcal{R}$ algebraisch ganz, so gibt es das Minimalpolynom von a in $\mathcal{R}[X]$.

2.2.5 Determinante und Adjunkte von Matrizen

Es sei an die Definition der Determinante einer Matrix als Polynom in den Einträgen der Matrix erinnert.

Definition 2.2.16 *Die* Determinante der Ordnung $n \in \mathbb{N}$ *ist ein Polynom*

$$\det \in \mathbb{Z}[X_{i,j} : i,j = 1,\ldots,n]$$

in n^2 *Variablen, definiert als*

$$\det := \sum_{\sigma\in\mathfrak{S}^n} sign(\sigma)\ X_{1,\sigma(1)} \cdots X_{n,\sigma(n)}\ .$$

Dabei ist $\mathfrak{S}^n$ *die Gruppe der Permutationen von* $\{1,\ldots,n\}$ *und* $sign(\sigma)$ *das Vorzeichen der Permutation* σ*, welches* 1 *ist, wenn* σ *durch eine gerade Anzahl an Vertauschungen erzeugt werden kann, sonst* $sign(\sigma) = -1$.

Man ordnet die Variablen in Form einer quadratischen $n \times n$–Matrix $\mathfrak{X}$ an, $X_{i,j}$ ist dann die Variable in Zeile i und Spalte j. Ist $\mathcal{R}$ ein beliebiger Ring und $\mathfrak{A} = \{A_{i,j}\}_{i,j=1,\ldots,n}$ eine quadratische $n \times n$–Matrix mit Einträgen aus $\mathcal{R}$, so kann das Determinantenpolynom in $\mathfrak{A}$ zu $\det(\mathfrak{A}) \in \mathcal{R}$ ausgewertet werden. Ist $\mathfrak{A}$ eine Matrix mit Einträgen aus einem Körper $\Bbbk$, so gilt $\det(\mathfrak{A}) = 0$ genau dann, wenn die Spalten von $\mathfrak{A}$ linear abhängig sind.

Aus einer Matrix $\mathfrak{A}$ mit Einträgen in $\mathcal{R}$ kann eine Matrix mit Einträgen in $\mathcal{R}[X]$ als $X\,I_n - \mathfrak{A} = \{\delta_{i,j}X - A_{i,j}\}_{i,j=1,\ldots,n}$ konstruiert werden. Deren Determinante $\chi_A(X) := \det(X\,I_n - \mathfrak{A})$ ist ein normiertes Polynom in X, dessen Koeffizienten Polynome in den Einträgen von $\mathfrak{A}$ sind,

$$\chi_A(X) = X^n + c_{n-1}X^{n-1} + \cdots + c_1X + c_0\ .$$

Insbesondere sind $-c_{n-1} = \mathrm{spur}(\mathfrak{A})$ die Spur der Matrix A und $(-1)^n c_0 = \det(\mathfrak{A})$ die Determinante der Matrix. Dieses Polynom kann auf dem von $\mathfrak{A}$ erzeugten kommutativen Matrixring ausgewertet werden, nach dem *Satz von Cayley–Hamilton* gilt $\chi_A(A) = 0$.

Betrachten wir die Matrix $\mathfrak{X}$ der Variablen $X_{1,1},\ldots,X_{n,n}$ als n–Tupel $\mathfrak{X} = (\mathbf{X}_1,\ldots,\mathbf{X}_n)$ von Spaltenvektoren $\mathbf{X}_k = (X_{1,k},\ldots,X_{n,k})^t$, so ist das Polynom det in jeder der Spalten linear. Es gilt also z.B.

$$\begin{aligned}\det = X_{1,1}\det(\mathbf{e}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n) + X_{2,1}\det(\mathbf{e}_2,\mathbf{X}_2,\ldots,\mathbf{X}_n) + \ldots \\ \cdots + X_{n,1}\det(\mathbf{e}_n,\mathbf{X}_2,\ldots,\mathbf{X}_n)\ ,\end{aligned} \quad (2.1)$$

wobei $\mathbf{e}_1,\ldots,\mathbf{e}_n$ die kanonischen Einheitsvektoren sind, die nur 0 und 1 als Komponenten haben, und zwar hat $\mathbf{e}_k$ eine Eins an der Stelle k, $k = 1,\ldots,n$, sonst Nullen. Mit anderen

Worten, $I_n = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ ist die Einheitsmatrix der Ordnung n. Nach dem *Laplaceschen Entwicklungssatz* sind die in dieser Formel auftretenden Determinanten nur von den Teilmatrizen der Ordnung $n-1$ von X abhängig, die durch Streichen der ersten Spalte und einer Zeile entsteht. Solche Determinanten heißen *Minoren der Ordnung* $n-1$.

Definition 2.2.17 *Seien* $\mathbf{A}_1, \ldots, \mathbf{A}_n \in \mathcal{R}^n$ *Spaltenvektoren mit Einträgen aus einem Ring* $\mathcal{R}$ *und* $\mathfrak{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_n)$ *die aus diesen zusammengesetzte* $n \times n$*–Matrix. Die* adjunkte Matrix *oder einfach* Adjunkte $\mathfrak{A}^\#$ *von* $\mathfrak{A}$ *ist diejenige* $n \times n$*–Matrix, deren Einträge die Minoren der Ordnung* $n-1$ *von* $\mathfrak{A}$ *sind,*

$$\mathfrak{A}^\# = \{(\mathfrak{A}^\#)_{i,j}\}_{i,j=1,\ldots,n} \quad \textit{mit} \quad (\mathfrak{A}^\#)_{i,j} := \det\left(\mathbf{A}_1, \ldots, \mathbf{A}_{i-1}, \mathbf{e}_j, \mathbf{A}_{i+1}, \ldots, \mathbf{A}_n\right) .$$

Man überzeugt sich leicht, dass analog zu Gleichung (2.1)

$$\mathfrak{A}^\# \cdot \mathfrak{A} = \mathfrak{A} \cdot \mathfrak{A}^\# = \det(\mathfrak{A})\, I_n \tag{2.2}$$

gilt. Jedes Gleichungssystem der Art $\mathfrak{A} \cdot \mathbf{x} = \mathbf{b}$ mit einer quadratischen $n \times n$–Matrix $\mathfrak{A}$ und Spaltenvektoren $\mathbf{b}$ und $\mathbf{x}$ der Dimension n kann also zu $\det(\mathfrak{A})\mathbf{x} = \mathfrak{A}^\# \cdot \mathbf{b}$ umgeformt werden. Ist $\det(\mathfrak{A}) \neq 0$ und $\mathcal{R}$ ein Integritätsbereich, so ist diese Umformung eine Äquivalenz. Ist $\det(\mathfrak{A})$ darüberhinaus im Ring $\mathcal{R}$ invertierbar, so kann das Gleichungssystem nach

$$\mathbf{x} = \det(\mathfrak{A})^{-1}\mathfrak{A}^\# \cdot \mathbf{b} \tag{2.3a}$$

gelöst werden, dies entspricht der *Cramerschen Regel*. Mit $\tilde{\mathbf{x}} := \mathfrak{A}^\# \cdot \mathbf{b}$ gilt immer

$$\mathfrak{A} \cdot \tilde{\mathbf{x}} = \det(\mathfrak{A})\, \mathbf{b} . \tag{2.3b}$$

Lemma 2.2.18 (Samuelson–Formel) *Seien* $\mathfrak{A} \in \mathcal{R}^{n\times n}$ *eine Matrix mit Einträgen aus einem Ring* $\mathcal{R}$*, sowie* $\mathbf{b}, \mathbf{c} \in \mathcal{R}^n$ *Spaltenvektoren und* $d \in \mathcal{R}$*. Dann gilt für die Determinante der aus diesen Bestandteilen zusammengesetzte* $(n+1) \times (n+1)$*–Matrix:*

$$\det \begin{pmatrix} \mathfrak{A} & \mathbf{c} \\ \mathbf{b}^t & d \end{pmatrix} = d\, \det \mathfrak{A} - \mathbf{b}^t \mathfrak{A}^\# \mathbf{c} .$$

Beweis: Durch Laplace–Entwicklung der erweiterten Matrix zuerst nach der letzten Zeile und dann nach der letzten Spalte erhalten wir

$$\begin{aligned} \det \begin{pmatrix} \mathfrak{A} & \mathbf{c} \\ \mathbf{b}^t & d \end{pmatrix} &= \sum_{j=1}^{n} b_j \det \begin{pmatrix} \mathfrak{A} & \mathbf{c} \\ \mathbf{e}_j^t & 0 \end{pmatrix} + d \det \begin{pmatrix} \mathfrak{A} & \mathbf{c} \\ 0 & 1 \end{pmatrix} \\ &= \sum_{j=1}^{n} \sum_{k=1}^{n} b_j c_k \det \begin{pmatrix} \mathfrak{A} & \mathbf{e}_k \\ \mathbf{e}_j^t & 0 \end{pmatrix} + d \det \mathfrak{A} \end{aligned}$$

Das Vertauschen der Spalte j mit der Spalte $n+1$ ändert das Vorzeichen der Determinante, und wir erhalten

$$\det \begin{pmatrix} \mathfrak{A} & \mathbf{e}_k \\ \mathbf{e}_j^t & 0 \end{pmatrix} = -\det \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{e}_k & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n & 0 \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} = -(\mathfrak{A}^\#)_{j,k} .$$

Durch Einsetzen in die Entwicklungsformel ergibt sich die Samuelson–Formel:

$$\det \begin{pmatrix} \mathfrak{A} & \mathbf{c} \\ \mathbf{b}^t & d \end{pmatrix} = -\sum_{j=1}^{n}\sum_{k=1}^{n} b_j c_k \, (\mathfrak{A}^{\#})_{j,k} + d \, \det \mathfrak{A} = d \, \det \mathfrak{A} - \mathbf{b}^t \mathfrak{A}^{\#} \mathbf{c}$$

□

Mit dem charakteristischen Polynom einer Matrix $\mathfrak{A}$ gilt

$$0 = \chi_{\mathfrak{A}}(\mathfrak{A}) = \mathfrak{A}^n + c_{n-1}\mathfrak{A}^{n-1} + \cdots + c_1 \mathfrak{A} + (-1)^n \det(\mathfrak{A}) \, ,$$

woraus man für die Adjunkte $\mathfrak{A}^{\#} = (-1)^{n-1}(\mathfrak{A}^{n-1} + c_{n-1}\mathfrak{A}^{n-2} + \cdots + c_1)$ abliest. Kennt man also das charakteristische Polynom einer Matrix und existiert das inverse Element zur Determinante, so kann ein lineares Gleichungssystem durch Anwenden dieser Formel durch $n-1$ Matrix–Vektor–Produkte und n skalare Multiplikationen der entstehenden Vektoren bestimmt werden.

Weiter kann diese Formel in die Samuelson–Formel eingesetzt werden, um das charakteristische Polynom der erweiterten Matrix zu bestimmen.

Lemma 2.2.19 (Schritt des Berkowitz–Algorithmus [Ber84, Abd96])
Sei $\mathfrak{A}$ eine $n \times n$–Matrix mit Einträgen aus einem Ring $\mathcal{R}$ und $\chi_A = X^n + c_{n-1}X^{n-1} + \cdots + c_0$ ihr charakteristisches Polynom. Sei mit den Spaltenvektoren $b, c \in \mathcal{R}^n$ und $d \in \mathcal{R}$ die $(n+1) \times (n+1)$–Matrix

$$\tilde{\mathfrak{A}} := \begin{pmatrix} \mathfrak{A} & \mathbf{c} \\ \mathbf{b}^t & d \end{pmatrix}$$

gebildet. Dann ergeben sich die Koeffizienten des charakteristischen Polynoms $\chi_{\tilde{\mathfrak{A}}}(X) = X^{n+1} + \tilde{c}_n X^n + \cdots + \tilde{c}_0$ zu

$$\tilde{c}_n = c_{n-1} - d, \quad \tilde{c}_0 = -d\, c_0 - \sum_{m=0}^{n-1} c_{m+1} \mathbf{b}^t A^m \mathbf{c}$$

$$\text{und} \quad \tilde{c}_k = c_{k-1} - d\, c_k - \sum_{m=0}^{n-1-k} c_{k+m+1} \mathbf{b}^t A^m \mathbf{c} \quad \text{für} \quad k = 1, \ldots, n-1 \, .$$

Beweis: Für das charakteristische Polynom der Matrix $\mathfrak{A} - Y\, I_n$ erhalten wir in $\mathcal{R}[X, Y]$

$$\chi_{(\mathfrak{A} - Y\, I_n)}(X) = \det(X\, I - (\mathfrak{A} - Y\, I_n)) = \chi_{\mathfrak{A}}(X + Y) \, .$$

Daraus erhält man die Adjunkte von $Y\, I_n - \mathfrak{A}$ zu

$$(Y\, I_n - \mathfrak{A})^{\#} = \sum_{k=0}^{n-1} \left(\sum_{m=0}^{n-1-k} c_{k+m+1} A^m \right) Y^k \, .$$

Einsetzen in die Samuelson–Formel liefert also

$$\begin{aligned} \det \begin{pmatrix} Y\, I_n - \mathfrak{A} & -\mathbf{c} \\ -\mathbf{b}^t & Y - d \end{pmatrix} &= (Y - d) \, \det(Y\, I_n - \mathfrak{A}) - \mathbf{b}^t (Y\, I_n - \mathfrak{A})^{\#} \mathbf{c} \\ &= \sum_{k=0}^{n} c_k (Y^{k+1} - dY^k) - \sum_{k=0}^{n}\sum_{k=0}^{n-1} \left(\sum_{m=0}^{n-1-k} c_{k+m+1} \mathbf{b}^t A^m \mathbf{c} \right) Y^k \, . \end{aligned}$$

Durch Zusammenfassen der Koeffizienten gleichen Grades ergibt sich die Behauptung. □

Der *Berkowitz–Algorithmus* (s. [Ber84, Abd96]) ergibt sich, indem man für eine $n \times n$–Matrix diesen Schritt auf die linken oberen $k \times k$–Untermatrizen anwendet, angefangen von der 1×1–Matrix des linken oberen Elements, deren charakteristisches Polynom trivial ist. Dieser Algorithmus ist divisionsfrei. Durch geschicktes Zusammenfassen der Zwischenergebnisse kann erreicht werden, dass zur Bestimmung des charakteristischen Polynoms einer $n \times n$–Matrix nur $O(n^4)$ Ringmultiplikationen benötigt werden.

2.2.6 Resultante und Diskriminante

Sind zwei Polynome $a, b \in \mathbb{Q}[X, Y]$ in zwei Variablen gegeben, so kann die Varietät $V(a, b) \in \mathbb{C}^2$ als Schnittmenge der Kurven $V(a)$ und $V(b)$ aufgefasst werden. Als solche kann $V(a, b)$ sowohl Teilkurven enthalten, die den Varietäten $V(a)$ und $V(b)$ gemeinsam sind, als auch „echte“ Schnittpunkte. Die beiden gemeinsamen Kurven entsprechen gemeinsamen Faktoren der Polynome a und b. Um diese Polynome auf gemeinsame Faktoren zu untersuchen, betrachten wir sie als univariate Polynome in der Variablen Y über dem Grundring $\mathcal{R} := \mathbb{Q}[X]$. Dieser ist ein faktorieller Integritätsbereich.

Seien also $\mathcal{R}$ ein faktorieller Integritätsbereich und angenommen, dass $a, b \in \mathcal{R}[Y]$ einen gemeinsamen Faktor besitzen, d.h. es gebe weitere univariate Polynome $f, g, h \in \mathcal{R}[Y]$ mit $a := fg$, $b := fh$ und $\deg f > 0$. Aus der trivialen Identität

$$0 = ab - ba = f(ah - bg)$$

kann, da $\mathcal{R}[Y]$ ein Integritätsbereich ist, der gemeinsame Faktor f gekürzt werden, es ergibt sich die Identität $0 = ah - bg$.

Seien weiter $m := \deg a$ und $n := \deg b$, d.h.

$$a = a_0 + a_1 Y + \cdots + a_m Y^m \text{ und } b = b_0 + b_1 Y + \cdots + b_n Y^n \,.$$

Da $\deg f > 0$ und $\mathcal{R}$ ein Integritätsbereich ist, gilt auch $\deg g = \deg a - \deg f < \deg a$ und analog $\deg h < \deg b$.

Bezeichnen wir mit $\mathcal{R}[Y]_d$ die Menge aller Polynome vom Grad kleiner d, so hat die $\mathcal{R}$–lineare Abbildung

$$S : \mathcal{R}[Y]_n \times \mathcal{R}[Y]_m \to \mathcal{R}[Y]_{m+n}, \quad (u, v) \mapsto au + bv$$

die nichttriviale Nullstelle $(h, -g)$. Wir können die Koeffizientenfolgen der Polynome in $\mathcal{R}[Y]_d$ als Spaltenvektoren der Länge d darstellen. Die Koeffizienten zum niedrigsten Grad seien als oberste Komponenten des Spaltenvektors aufgeführt. Wir können in $w := S(u, v)$ die Koeffizientenfolgen von u und v zu einem Spaltenvektor

$$(u, v)^t := (u_0, \ldots, u_{n-1}, v_0, \ldots, v_{m-1})^t$$

der Länge $m+n$ zusammenfassen. Da $\deg w < m+n$ nach Konstruktion gilt, also $w \in \mathcal{R}[Y]_{m+n}$, können wir S eine quadratische $(m+n) \times (m+n)$–Matrix $Syl(a,b)$ zuordnen, diese wird *Sylvester-Matrix* genannt:

$$Syl(a,b) := \begin{pmatrix} a_0 & & & & b_0 & & & \\ a_1 & a_0 & & & b_1 & b_0 & & \\ \vdots & a_1 & \ddots & & \vdots & b_1 & \ddots & \\ \vdots & \vdots & & a_0 & \vdots & \vdots & \ddots & b_0 \\ a_m & \vdots & & a_1 & b_n & \vdots & & b_1 \\ & a_m & & \vdots & & b_n & & \vdots \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & a_m & & & & b_n \end{pmatrix} . \tag{2.4}$$

Da $S(h, -g) = 0$ ist, gilt auch für das Matrix–Vektor–Produkt $Syl(a,b)\,(g, -h)^t = 0$. Nach Multiplikation mit der Adjunkten $Syl(a,b)^{\#}$ ergibt sich daraus

$$\det Syl(a,b)\,(g, -h)^t = 0\,.$$

Es gibt von Null verschiedene Koeffizienten in g und h. Da $\mathcal{R}$ als Integritätsbereich vorausgesetzt war, ist $\det Syl(a,b) = 0$.

Definition 2.2.20 *Die Determinante der Sylvester–Matrix von $a, b \in \mathcal{R}[X]$ wird Resultante genannt und als* $\mathrm{Res}(a,b) := \det Syl(a,b)$ *notiert.*

Ist $\mathrm{Res}(a,b)$ von Null verschieden, so kann das Gleichungssystem $au + bv = \mathrm{Res}(a,b)$ gelöst werden, der Spaltenvektor $(u,v)^t$ der Koeffizienten von u und v ergibt sich nach der Cramerschen Regel (2.3b) als erste Spalte der Adjunkten der Sylvestermatrix. Die Koeffizienten von u und v sind also Polynome des Grades $m+n-1$ in den Koeffizienten von a und b.

Satz 2.2.21 (s. [GG99], Satz 6.14) *Sei $\mathcal{R}$ ein faktorieller Integritätsbereich und seien $a, b \in \mathcal{R}[Y]$ univariate Polynome. Dann gilt* $\mathrm{Res}(a,b) = 0$ *genau dann, wenn es einen gemeinsamen Faktor von a und b gibt.*

Ist $\mathcal{R}$ ein Körper, so ist dies eine einfache Anwendung linearer Algebra. Im allgemeinen Fall wird wieder die Strategie des Nachweises der Behauptung über dem Quotientenkörper $\mathcal{K} := Quot(\mathcal{R})$ und der nachfolgenden Elimination von Hauptnennern und Abgleich von Inhalten der Polynome angewandt.

Beispiel: In Fortführung der einleitenden Betrachtung gibt es für $a, b \in \mathbb{Q}[X][Y]$ also einen gemeinsamen Faktor, wenn $\mathrm{Res}_Y(a,b) = 0$ gilt. Ist $f \in \mathbb{Q}[X][Y]$ der größte gemeinsame Teiler und $g, h \in \mathbb{Q}[X][Y]$ mit $a = fg$ und $b = fh$, so gibt es weitere Polynome $u, v \in \mathbb{Q}[X][Y]$ mit $w := gu + hv = \mathrm{Res}(g,h) \in \mathbb{Q}[X]$. Ist $\xi \in \mathbb{C}$ eine Nullstelle von w, so haben entweder $g_\xi := g(\xi, Y)$, $h_\xi := h(\xi, Y) \in \mathbb{C}[Y]$ gemeinsame Nullstellen oder in beiden Spezialisierungen verschwindet der jeweils führende Koeffizient von g_ξ und h_ξ, da dann die letzte Zeile der Sylvester–Matrix konstant Null ist. Der zweite Fall entspricht, in einem angepassten theoretischen Rahmen, einer gemeinsamen Nullstelle in (ξ, ∞).

Seien $a, b \in \mathbb{Q}[X][Y]$ als in Y normiert vorausgesetzt, und folglich auch f normiert, so dass bei Spezialisierung in X der Grad immer erhalten bleibt. $V(a,b)$ besteht dann aus der Kurve $V(f)$ und allen Punkten

$(\xi, \eta) \in \mathbb{C}^2$, in welchen ξ eine Nullstelle von $w = \mathrm{Res}(g, h)$ und η eine Nullstelle des – nicht konstanten – größten gemeinsamen Teilers von g_ξ und h_ξ sind.

Seien $\mathcal{R}$ ein Integritätsbereich und $a \in \mathcal{R}[X]$ ein Polynom, welches einen zweifach auftretendem Faktor besitzt, d.h. $a = f^2 b$ mit $f, b \in \mathcal{R}[X]$. Für univariate Polynome kann die Ableitung auf rein algebraische Weise definiert werden, ist $a = a_0 + a_1 X + \cdots + a_m X^m \in \mathcal{R}[X]$, so ist dessen Ableitung gerade $a' = a_1 + 2a_2 X + \cdots + m a_m X^{m-1}$. Es gelten die üblichen Rechenregeln der Differentiation, insbesondere hat $a = f^2 b$ nach der Leibnizregel die Ableitung $a' = (p' f + 2pf')f$, d.h. f ist ein gemeinsamer Faktor von a und a'.

Definition 2.2.22 *Sei $\mathcal{R}$ ein Ring und $a = a_0 + a_1 X + \cdots + a_m X^m \in \mathcal{R}[X]$ ein univariates Polynom. Die* Diskriminante *von a ist die um den führenden Koeffizienten a_m von a reduzierte Resultante von a und der Ableitung $a' = a_1 + 2a_2 X + \cdots + m a_m X^{m-1}$. Sie wird als* $\mathrm{Disk}(a)$ *notiert, es gilt* $a_m \,\mathrm{Disk}(a) = \mathrm{Res}(a, a')$ *notiert.*

Dies ist wohldefiniert, da in der letzten Zeile der Sylvestermatrix, die den höchsten Koeffizienten der Polynome zugeordnet ist, neben Nulleinträgen nur die Einträge a_m und $m a_m$ vorkommen. Die Resultante von a und a' ist also immer ein Vielfaches von a_m.

Satz 2.2.23 *Seien $\mathcal{R}$ ein faktorieller Integritätsbereich und $a \in \mathcal{R}[X]$ ein univariates Polynom. Gilt* $\mathrm{Disk}(a) = 0$*, so gibt es einen nichttrivialen Faktor f von a, so dass auch f^2 ein Faktor von a ist.*

Beweis: Da $\mathcal{R}$ keine Nullteiler enthält, und der führende Koeffizient a_m von Null verschieden ist, ist die Diskriminante von a genau dann Null, wenn auch die Resultante verschwindet. Ist also ein Faktor von a auch quadratisch in a enthalten, so verschwindet die Diskriminante.

Es verbleibt der Fall zu betrachten, in welchem aus $\mathrm{Disk}(a) = 0$ das Auftreten quadratischer Faktoren zu folgern ist. Nach Satz 2.2.21 gibt es einen nichtkonstanten größten gemeinsamen Teiler $p \in \mathcal{R}[X]$ von a und a'. Sei $q \in \mathcal{R}[X]$ der verbleibende Faktor in $a = pq$. Nun gilt aber auch $a' = p' q + p q'$. Da p ein Teiler von a' ist, muss p auch das Produkt $p' q$ teilen. p kann nicht vollständig in p' enthalten sein, da $\deg p > \deg p'$ gilt. Also muss es einen nichtkonstanten Faktor $f \in \mathcal{R}[X]$ von p geben, der auch q teilt. Somit ist f^2 ein Teiler von a. □

Ist $\Bbbk$ algebraisch abgeschlossen, so hat jedes $a \in \Bbbk[X]$ mit $\mathrm{Disk}(a) = 0$ mehrfache Nullstellen in $\Bbbk$, d.h. in der Zerlegung von a in Linearfaktoren gibt es mehrfach bzw. mit einer Potenz größer als Eins auftretende Faktoren.

Ist $\mathcal{R}$ ein faktorieller Integritätsbereich, so kann jedes Polynom $a \in \mathcal{R}[X]$ in Primfaktoren zerlegt werden. Fassen wir Primfaktoren gleicher Vielfachheit zusammen, so erhalten wir eine Darstellung

$$a = f_1 f_2^2 \cdots f_M^M \,, \tag{2.5}$$

in welcher jeder der Faktoren $f_1, \ldots, f_M \in \mathcal{R}[X]$ quadratfrei ist. Bestimmen wir in dieser Darstellung die Ableitung von a, so erhalten wir

$$a' = (f_1' f_2 \cdots f_M + 2 f_1 f_2' \cdots f_M + \cdots + M f_1 f_2 \cdots f_M') f_2 f_3^2 \cdots f_M^{M-1} \,.$$

Der größte gemeinsame Teiler von a und a' ist also $a_1 = f_2 f_3^2 \cdots f_M^{M-1}$, der verbleibende Faktor nach Reduktion von a um a_1 ist $b_1 = f_1 f_2 \ldots f_M$.

Die oben angegebene Faktorisierung von a erfolgte auf nichtkonstruktive Weise. Es läßt sich aber eine konstruktive Bestimmung dieser Faktorisierung ablesen. Man setze $a_0 := a$ und bestimme rekursiv

$$a_k := ggT(a_{k-1}, a'_{k-1}), \qquad b_k := a_{k-1}/a_k, \qquad k = 1, \ldots, M$$

bis zu einem M mit $\deg a_M = 1$. Dabei erhält man eine Folge von Polynomen $b_k = f_k f_{k+1} \ldots f_M$, aus welcher sich die quadratfreien Faktoren $f_1, \ldots, f_M$ als $f_k := b_k/b_{k+1}$ bestimmen lassen.

2.2.7 Arithmetik algebraischer Elemente

Mit Hilfe der Resultante kann man nun die arithmetischen Operationen indirekt gegebener algebraischer Elemente definieren. Als Beispiel können die algebraischen Zahlen in $\mathbb{R}$ dienen. Diese können indirekt durch ein Polynom mit ganzzahligen Koeffizienten und ein eine Nullstelle isolierendes Intervall angegeben werden. Dieses einschließende Intervall kann mit Verfahren zur Nullstellenbestimmung wie z.B. dem Sekantenverfahren bei Bedarf beliebig verkleinert werden. Die Verknüpfung der Intervalle unter Addition und Multiplikation ist trivial, es verbleibt, die Verknüpfung der Polynome zu bestimmen.

Satz 2.2.24 *Sei $\mathcal{R}$ ein faktorieller Integritätsbereich und $\mathcal{A}$ eine $\mathcal{R}$–Algebra. Seien $a, b \in \mathcal{A}$ algebraisch über $\mathcal{R}$. Dann sind auch $a + b$, ab und jedes Vielfache ra mit $r \in \mathcal{R}$ algebraisch über $\mathcal{R}$. Sind a, b algebraisch ganz, so auch $a + b$, ab und ra.*

Beweis: Als erstes sei bemerkt, dass jedes Element $r \in \mathcal{R}$ algebraisch ganz über $\mathcal{R}$ ist mit Minimalpolynom $q(X) := X - r$. Es genügt also, die „echten“ Operationen zu untersuchen. Seien $p, q \in \mathcal{R}[X]$ Polynome mit $p(a) = 0$ und $q(b) = 0$. Sei $m := \deg p$ und $n := \deg q$.

Für die Summe $a + b$ betrachten wir die Polynome $p(Y), q(X - Y) \in \mathcal{R}[X][Y]$ und deren Resultante $w := \mathrm{Res}_Y(p(Y), q(X - Y)) \in \mathcal{R}[X]$. Die höchste Potenz von X in $q(X - Y)$ ist das Monom $q_n X^n$, dieses ist bzgl. Y konstant. Das Monom höchsten Grades von $w \in \mathcal{R}[X]$ ergibt sich daher aus dem Produkt der Diagonalelemente der Sylvestermatrix als

$$p_m^n (q_n X^n)^m = p_m^n q_n^m X^{mn} .$$

Der Grad von w ist also $\deg w = mn$. Sind p und q normiert, so ist auch w normiert. Es gibt weiter Polynome $u, v \in \mathcal{R}[X, Y]$ mit $w(X) = p(Y)u(X, Y) + q(X - Y)v(X, Y)$. Werten wir diese Identität in $X = a + b$, $Y = a$ aus und beachten $w \in \mathcal{R}[X]$, so folgt

$$w(a + b) = p(a)u(a + b, a) + q(b)v(a + b, a) = 0 ,$$

$a + b$ ist somit algebraisch. Sind die Summanden algebraisch ganz über $\mathcal{R}$, so gilt dies auch für die Summe.

Um das Produkt ab zu untersuchen, betrachten wir zusätzlich zum Polynom $p(Y)$ das Polynom $\tilde{q}(X, Y) := q_0 Y^n + q_1 Y^{n-1} X + \cdots + q_n X^n$. Das Monom $q_n X^n$ höchsten Grades in X

von $\tilde{p}$ ist wieder in Y konstant, somit ergibt sich wie oben das Monom höchsten Grades von $w \in \mathcal{R}[X]$ aus dem Produkt der Diagonalelemente der Sylvestermatrix zu $p_m^n q_n^m X^{mn}$. Da dies nach Konstruktion von Null verschieden ist, sind die Polynome $p, \tilde{q}$ teilerfremd. Es gibt daher Polynome $u, v \in \mathcal{R}[X, Y]$ mit $w = qu + \tilde{p}v$. Es gilt weiter $\tilde{p}(bY, Y) = Y^n p(b) = 0$, unter Auswertung in $X = ab$, $Y = a$ erhalten wir also

$$w(ab) = q(a)u(ab, a) + \tilde{p}(ab, a)v(ab, a) = 0\,.$$

ab ist also algebraisch über $\mathcal{R}$. Sind die Faktoren algebraisch ganz, so gilt dies auch für das Produkt. Denn mit $p_m = q_n = 1$ ist auch das Polynom w normiert. □

Korollar 2.2.25 *Seien $\mathcal{R}$ ein faktorieller Integritätsbereich und $\mathcal{A} := \mathcal{R}[a_1, \ldots, a_n]$ eine endlich erzeugte $\mathcal{R}$–Algebra. Sind die erzeugenden Elemente $a_1, \ldots, a_n$ algebraisch über $\mathcal{R}$, so ist $\mathcal{A}$ algebraisch über $\mathcal{R}$, sind $a_1, \ldots, a_n$ algebraisch ganz, so auch $\mathcal{A}$.*

Beweis: Jedes Polynom in $\mathcal{R}[X_1, \ldots, X_n]$ ist eine endliche Summe von Produkten mit endlich vielen Faktoren. Daher kann Satz 2.2.24 rekursiv angewandt werden, um zu jedem Element $a \in \mathcal{A}$ ein Polynom bzw. ein normiertes Polynom $p \in \mathcal{R}[Y]$ zu konstruieren, welches in a verschwindet. □

Das nach dem vorhergehenden Beweis konstruierte Vielfache des Minimalpolynoms hat einen exponentiell in der Anzahl der notwendigen Rechenoperationen wachsenden Grad. Dies kann jedoch verbessert werden.

Lemma 2.2.26 *Seien $\mathcal{R}$ ein Integritätsbereich, $q_1, \ldots, q_n \in \mathcal{R}[X]$ normierte univariate Polynome und*

$$\mathcal{A} := \mathcal{R}[X_1, \ldots, X_n] \big/ \langle\, q_1(X_1), \ldots, q p_n(X_n) \,\rangle\,.$$

Dann hat jedes Element von $\mathcal{A}$ ein normiertes Minimalpolynom mit einem durch $D := \deg p_1 \ldots \deg p_n$ beschränkten Grad.

Beweis: Sei im Gitter $\mathbb{N}^n$ der durch die Grade der Polynome $q_1, \ldots, q_n$ rechteckig beschränkte Bereich

$$K := \{\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n : \alpha_1 < \deg q_1, \ldots, \alpha_n < \deg q_n\}$$

fixiert. Man überlegt sich leicht, dass $\mathcal{A}$ von den Restklassen $e_\alpha := X^\alpha + I$ der Monome mit Multiindex $\alpha \in K$ aufgespannt wird. Sei $a \in \mathcal{A}$ ein beliebiges Element. Dann gibt es für jeden Multiindex $\alpha \in K$ Koeffizienten $A_{\alpha,\beta} \in \mathcal{R}$, $\beta \in K$, mit $e_\alpha a = \sum_{\beta \in K} A_{\alpha,\beta} e_\beta$. Bringt man die Multiindizes von K in irgendeine Reihenfolge und betrachtet dementsprechend $A := (A_{\alpha,\beta})_{\alpha,\beta \in K}$ als Matrix mit Einträgen aus $\mathcal{R}$, so gilt für den Spaltenvektor $\{e_\beta\}_{\beta \in K}$

$$(A - aI)\{e_\beta\}_{\beta \in K} = \left\{ \sum_{\beta \in K} (A_{\alpha,\beta} - a\delta_{\alpha,\beta}) e_\beta \right\} = 0\,.$$

$I = (\delta_{\alpha,\beta})_{\alpha,\beta \in K}$ bezeichnet dabei die Einheitsmatrix mit dem Kronecker–Symbol $\delta_{\alpha,\beta}$, es gelten $\delta_{\alpha,\alpha} = 1$ und $\delta_{\alpha,\beta} = 0$ für $\alpha, \beta \in K$, $\alpha \neq \beta$. Das charakteristische Polynom χ_A von A hat somit

in a eine Nullstelle. Nach Konstruktion ist χ_A normiert, damit auch das Minimalpolynom von a, welches ein Faktor von χ_A ist. Der Grad von χ_A, der eine Schranke für den Grad des Minimalpolynoms darstellt, stimmt mit der Mächtigkeit von K überein, diese beträgt gerade $D = \deg p_1 \ldots \deg p_n$. □

Korollar 2.2.27 *Seien $\mathcal{R}$ ein faktorieller Integritätsbereich und $\mathcal{A} = \mathcal{R}[a_1, \ldots, a_n]$ eine endlich erzeugte $\mathcal{R}$–Algebra. Sind die erzeugenden Elemente $a_1, \ldots, a_n \in \mathcal{A}$ algebraisch ganz über $\mathcal{R}$, so gibt es eine endliche Teilmenge von $\mathcal{A}$, welche $\mathcal{A}$ in Form von Linearkombination mit Koeffizienten in $\mathcal{R}$ aufspannt.*

Seien $q_1, \ldots, q_n \in \mathcal{R}[X]$ die Minimalpolynome von $a_1, \ldots, a_n$ und $D := \deg p_1 \ldots \deg p_n$. Dann ist die Anzahl der Elemente in dieser aufspannenden Menge durch D beschränkt, jedes Element von RA hat ein Minimalpolynom mit durch D beschränktem Grad.

Beweis: Sei $\varphi : \mathcal{R}[\underline{X}] \to \mathcal{A}$ die Auswertungsabbildung des Polynomrings $\mathcal{R}[\underline{X}] = \Bbbk[X_1, \ldots, X_n]$ zum Punkt $(a_1, \ldots, a_n) \in \mathcal{A}^n$. Das von den Polynomen q_k erzeugte Ideal $I := \langle q_1(X_1), \ldots, q_n(X_n) \rangle$ wird unter φ auf die Null abgebildet, somit ist der $\mathcal{R}$–Algebrenmorphismums $\varphi_* : \mathcal{R}[\underline{X}]/_I \to \mathcal{A}$ wohldefiniert, $\mathcal{R}$–linear und weiterhin surjektiv. Man überlegt sich leicht, dass $\mathcal{R}[\underline{X}]/_I$ von den Restklassen der Monome X^α mit $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ und $\alpha_k < \deg q_k$ aufgespannt wird. $\mathcal{A}$ wird dementsprechend von den Bildern dieser Monome aufgespannt. Diese aufspannende Teilmenge von $\mathcal{A}$ hat also maximal $D := \deg q_1 \cdots \deg q_n$ Elemente.

Zu jedem Element $a \in \mathcal{A}$ gibt es ein Urbild $f \in \mathcal{R}[\underline{X}]/_I$. Nach dem vorhergehenden Lemma gibt es ein normiertes Minimalpolynom $q \in \mathcal{R}[X]$ von f, dessen Grad D nicht übersteigt. Dann gilt $q(a) = q(\varphi_*(f)) = \varphi_*(q(f)) = 0$, d.h. das Minimalpolynom von a ist ein Faktor von q. □

2.2.8 Parametrisierung à la Kronecker

Sind $\Bbbk$ ein Körper, $\overline{\Bbbk}$ ein algebraischer Abschluss von $\Bbbk$ und $V \subset \overline{\Bbbk}^n$ eine algebraische Varietät in Noether–Normalform der Dimension r, so ist der Koordinatenring $\Bbbk[V]$ eine $\Bbbk[X_1, \ldots, X_r]$–Algebra und die Restklassen der Variablen sind algebraisch ganz über $\Bbbk[X_1, \ldots, X_r]$. Sei D das Produkt der Grade der Minimalpolynome dieser Restklassen.

Die Werte jeder polynomialen Funktion $f \in \Bbbk[V]$ auf V sind somit Nullstellen von Polynomen aus $\Bbbk[X_1, \ldots, X_r][Y]$ vom maximalen Grad D. Betrachtet man $m \in \mathbb{N}$ Funktionen $f_1, \ldots, f_m \in \Bbbk[V]$ gleichzeitig, so erhält man bis zu D^m Kombinationen der Nullstellen der Minimalpolynome dieser Funktionen. Unter diesen möchte man die Kombinationen herausfinden, die auch tatsächlich als Wertetupel in Punkten von V auftreten.

Die Grundkonstruktion dazu betrachten wir in der allgemeinen Situation eines faktoriellen Integritätsbereichs $\mathcal{R}$, z.B. $\mathcal{R} = \Bbbk[X_1, \ldots, X_r]$ und einer endlich erzeugten $\mathcal{R}$–Algebra $\mathcal{A} =$

$\mathcal{R}[a_1,\ldots,a_m]$, z.B. $a_k = f_k \in \Bbbk[V]$, $k = 1,\ldots,m$, damit gilt $\mathcal{A} \subset \Bbbk[V]$. Seien die erzeugenden Elemente $a_1,\ldots,a_m$ algebraisch ganz über $\mathcal{R}$.

Dann sind diese Elemente auch algebraisch ganz über einem um m Parameter erweiterten Grundring $\mathcal{R}[\underline{T}] := \mathcal{R}[T_1,\ldots,T_m]$. Mit diesen Parametern kann die Linearkombination $U := T_1a_1 + \cdots + T_ma_m \in \mathcal{A}[\underline{T}]$ gebildet werden. Nach Korollar 2.2.25 ist auch U algebraisch ganz über $\mathcal{R}[\underline{T}]$, es gibt ein gradminimales normiertes Polynom $\tilde{Q} \in \mathcal{R}[\underline{T}][Y]$ mit $\tilde{Q}(\underline{T})(U) = 0$.

Sei $Q \in \mathcal{R}[\underline{T}][Y]$ der quadratfreie Teil von $\tilde{Q}$. Es gibt eine kleinste Potenz $M \in \mathbb{N}$, so dass $\tilde{Q}$ ein Teiler von Q^M ist. Nach Konstruktion gilt $Q(\underline{T})(U)^M = 0$, d.h. $Q(\underline{T})(U)$ ist für $M > 1$ ein nichttriviales nilpotentes Element der Algebra $\mathcal{A}$.

Wir können nun diese Identität partiell nach den freien Parametern $T_1,\ldots,T_m$ differenzieren. Die Ableitung der rechten Seite ergibt wieder Null, links erhalten wir

$$\frac{\partial}{\partial T_k}\left(Q(\underline{T})(U)^M\right) = MQ(\underline{T})(U)^{M-1}\left(\frac{\partial Q}{\partial T_k}(\underline{T})u(U) + \frac{\partial Q}{\partial Y}a_k\right) = 0\,.$$

Sei $\rho = \mathrm{Disk}(Q) \in \mathcal{R}[\underline{T}]$ die Diskriminante von Q nach Y. Da nach Konstruktion Q quadratfrei ist, gilt $\rho \neq 0$ und es gibt Polynome $A, B \in \mathcal{R}[\underline{T}][Y]$ mit $AQ' + BQ = \rho$. Mit Q^{M-1} multipliziert ergibt dies

$$(AQ^{M-1}Q')(\underline{T})(U) = Q^{M-1}(\underline{T})(U)\rho(\underline{T})\,.$$

Multiplizieren wir die partiellen Ableitungen mit $A(\underline{T})(U)$, so erhalten wir also

$$MQ(\underline{T})(U)^{M-1}\left(A(\underline{T})(U)\frac{\partial Q}{\partial T_k}(\underline{T})(U) + \rho(\underline{T})a_k\right) = 0\,.$$

Seien $W_k := -A\frac{\partial Q}{\partial T_k} \in \mathcal{R}[\underline{T}][Y]$.

Lemma 2.2.28 *Gilt in der oben konstruierten Situation schon $Q(\underline{T}, U) = 0$, d.h. $M = 1$, und gibt es ein Tupel $\lambda \in \mathcal{R}^n$, so dass $\rho(\lambda)$ in $\mathcal{R}$ invertierbar ist, so ist die $\mathcal{R}$–Algebra $\mathcal{A}$ isomorph zum Restklassenring $\mathcal{R}[Y]/\langle q \rangle$ mit $q(Y) := Q(\lambda)(Y) \in \mathcal{R}[Y]$.*

Beweis: Einerseits wird jede Klasse des Restklassenrings bei Einsetzten von $u := U(\lambda)$ für Y auf ein Element der Algebra $\mathcal{A}$ abgebildet. Da q das minimale Polynom für $u \in \mathcal{A}$ ist, wird nur die Restklasse der 0 auf das Nullelement der Algebra abgebildet, diese Auswertungsabbildung ist daher ein injektiver Homomorphismus zwischen beiden $\mathcal{R}$–Algebren $\mathcal{R}[Y]/\langle q \rangle$ und $\mathcal{A}$.

Seien andererseits univariate Polynome

$$w_k(Y) := \rho(\lambda)^{-1}W_k(\lambda)(Y) \in \mathcal{R}[Y], \qquad k = 1,\ldots,n,$$

definiert, mit diesen gilt $a_k = w_k(u)$. Jedes $b \in \mathcal{A}$ ist als Polynom $b = f(a_1,\ldots,a_n)$ in den Erzeugenden darstellbar. Das univariate Polynom $F(Y) := f(w_1(Y),\ldots,w_n(Y))$ definiert dann eine Restklasse in $\mathcal{R}[Y]/\langle q \rangle$. Nach Auswertung in u gilt $F(u) = f(w_1(u_\lambda),\ldots,w_n(u)) = b$. Die Auswertungsabbildung ist also auch surjektiv, somit ein Isomorphismus der $\mathcal{R}$–Algebren. □

Man kann jedes von Null verschiedene Element h eines Integritätsbereichs $\mathcal{R}$ „invertierbar machen“, indem man zu dem Teil $\mathcal{R}_h$ des Quotientenkörpers übergeht, dessen Brüche nur Potenzen von h im Nenner aufweisen. Diese Erweiterung des Rings $\mathcal{R}$ kann auch als Restklassenring $\mathcal{R}_h := \mathcal{R}[T]/\langle\, 1 - Th \,\rangle$ (Rabinowitsch–Trick) konstruiert werden.

Definition 2.2.29 $\mathcal{R}_h := \mathcal{R}[T]/\langle\, 1 - Th \,\rangle$ *wird* Lokalisierung *von $\mathcal{R}$ nach dem Element h genannt.*

Definition 2.2.30 *Seien $\mathcal{R}$ ein faktorieller Integritätsbereich und $\mathcal{A}$ eine über $\mathcal{R}$ algebraisch ganze Algebra. Ein $u \in \mathcal{A}$ mit Minimalpolynom $q \in \mathcal{R}[Y]$ und $\rho := \mathrm{Disk}(q)$ wird* primitives Element *der Algebra $\mathcal{A}$ genannt, wenn der Homomorphismus $\varphi_* : \mathcal{R}_\rho[Y]/\langle\, q(Y) \,\rangle \to \mathcal{A}_\rho$ der lokalisierten $\mathcal{R}_\rho$–Algebren, erzeugt durch $\varphi_*(Y + \langle\, q \,\rangle) := u$, bijektiv ist.*

Diese Grundkonstruktion kann nun auch auf eine algebraische Varietät in Noether–Normalform der Dimension r angewandt werden. Seien $\Bbbk$ ein Körper der Charakteristik 0, $\overline{\Bbbk}$ ein algebraischer Abschluss und V die algebraische Varietät. Nach Voraussetzung ist der Koordinatenring $\mathcal{A} := \Bbbk[V]$ eine $\mathcal{R}$–Algebra mit $\mathcal{R} = \Bbbk[X_1, \ldots, X_r]$. $\Bbbk[V]$ wird von den Restklassen der Variablen erzeugt, seien also $a_k := X + I(V)$, $k = 1, \ldots, n$. Nach Voraussetzung sind $a_1, \ldots, a_n$ algebraisch ganz über $\mathcal{R}$.

$\mathcal{A} = \Bbbk[V]$ enthält keine nilpotenten Elemente, da $I(V)$ radikal ist. Dies überträgt sich auch auf $\mathcal{A}[\,\underline{T}\,] = \mathcal{A}[T_1, \ldots, T_n]$. Die Linearform $U = T_1 a_1 + \cdots + T_n a_n$ besitzt also ein quadratfreies normiertes Minimalpolynom $Q \in \mathcal{R}[\,\underline{T}\,][Y]$. Damit ist dessen Diskriminante $\rho := \mathrm{Disk}_Y(Q) \in \mathcal{R}[\,\underline{T}\,] = \Bbbk[X_1, \ldots, X_r][\,\underline{T}\,]$ vom Nullpolynom verschieden.

Es gibt somit, da $\Bbbk$ unendlich viele Elemente besitzt, Tupel $\lambda = (\lambda_1, \ldots, \lambda_n) \in \Bbbk^n$, für welche nicht alle Koeffizienten der Monome in $\rho(\lambda) \in \Bbbk[X_1, \ldots, X_r]$ gleichzeitig verschwinden. Für dieses λ ist $u := \lambda_1 a_1 + \cdots + \lambda_n a_n$ mit dem Minimalpolynom $q := Q(\lambda) \in \mathcal{R}[Y]$ ein primitives Element von $\mathcal{A} = \Bbbk[V]$.

Definition 2.2.31 (Parametrisierung à la Kronecker, vgl. [Lec00]) *Seien $\Bbbk$ ein Körper der Charakteristik 0, $\overline{\Bbbk}$ ein algebraischer Abschluss und $V \subset \overline{\Bbbk}^n$ eine äquidimensionale algebraische Varietät der Dimension r.*

Eine Parametrisierung der Varietät à la Kronecker *besteht aus*

- *einer Noether–Normalisierung durch eine Matrix $A \in \Bbbk^{n\times n}$, d.h. $V^A := A^{-1}V$ befinde sich in Noether–Normalform der Dimension r,*
- *einem Tupel $\lambda = (\lambda_1, \ldots, \lambda_n) \in \Bbbk^n$, mit welchem die Restklasse u der Linearform $\lambda_1 X_1 + \cdots + u_n \lambda_n$ ein primitives Element im Koordinatenring $\Bbbk[V^A]$ ist,*
- *dem Minimalpolynom $q \in \Bbbk[X_1, \ldots, X_r][Y]$ von u mit Ableitung q' nach Y sowie*
- *Polynomen $v_1, \ldots, v_n \in \Bbbk[X_1, \ldots, X_r][Y]$, mit welchen $q'(u)x_1 = v_1(u), \ldots, q'(u)x_n = v_n(u)$ für die Klassen $x_1, \ldots, x_n \in \Bbbk[V^A]$ der Variablen $X_1, \ldots, X_n$ gilt.*

Ist wieder $\mathcal{R} := \Bbbk[X_1, \ldots, X_r]$, $\mathcal{A} := \Bbbk[V^A]$ und $\rho := \mathrm{Disk}_Y(q)$, so ist, da q quadratfrei ist, $\rho \neq 0$. $q'(u)$ ist in der Lokalisierung $\mathcal{A}_\rho$ invertierbar, denn es gibt Polynome $A, B \in \mathcal{R}[Y]$ mit $Aq' + Bq = \rho$, mit diesen ist $\rho^{-1}A(u)$ in $\mathcal{A}_\rho$ zu $q'(u)$ invers.

2.3 Kroneckers Methode

Die Aufgabe besteht darin, zu einem polynomialen Gleichungssystem mit Koeffizienten in $\mathbb{Q}$ eine endliche algebraische Erweiterung von $\mathbb{Q}$ zu konstruieren, in welcher sich schon alle Punkte der komplexen algebraischen Varietät, welche dieses Gleichungssystem beschreibt, darstellen lassen. Gleichzeitig soll auch eine Darstellung der Varietät in dieser Erweiterung gefunden werden. Die Methode nach Kronecker [Kro82] eliminiert nacheinander alle verwendeten Variablen, indem deren algebraische Abhängigkeit von den verbleibenden Variablen bestimmt wird. Wir skizzieren dieses Verfahren nach der Darstellung in [vW31].

Die Grundstruktur dieser Methode kann wie folgt angegeben werden. Seien $f_1, f_2 \in \mathcal{R}[X]$ zwei teilerfremde Polynome, wobei $\mathcal{R} = \mathbb{C}[\underline{T}] := \Bbbk[T_1, \ldots, T_m]$ ein Polynomring über dem Körper $\mathbb{C}$ ist, dessen Variable als Parameter aufgefasst werden. Nach Voraussetzung haben diese eine von Null verschiedene Resultante $\rho := \mathrm{Res}(f_1, f_2) \in \mathcal{R}$. Spezialisiert man die Parameter $\underline{T}$ zu einem Punkt $\tau \in \mathbb{C}^m$, in welchem die spezialisierte Resultante verschwindet, $\rho(\tau) = 0$, so haben die spezialisierten Polynome $f_1(\tau), f_2(\tau) \in \mathbb{C}[X]$ einen gemeinsamen Faktor, also mindestens eine gemeinsame Nullstelle $\eta \in \mathbb{C}$, sofern die Grade von f_1 bzw. f_2 mit denen von $f_1(\tau)$ bzw. $f_2(\tau)$ übereinstimmen. Sind die Koordinaten von $\tau \in \mathbb{C}^m$ algebraisch und die Koeffizienten von $f_1, f_2 \in \mathbb{Q}[\underline{T}][X]$ rational, so sind auch die gemeinsamen Nullstellen η algebraisch über $\mathbb{Q}$. Somit ist auch der Punkt $(\tau, \eta) \in \mathbb{C}^{m+1}$ algebraisch über $\mathbb{Q}$, welcher eine gemeinsame Nullstelle von f_1 und f_2 ist.

2.3.1 Vorbereitung mittels Koordinatenwechsel

Gegeben sei ein System $f_1, \ldots, f_s \in \mathbb{Q}[X_1, \ldots, X_n]$, von welchem zu bestimmen ist, ob es gemeinsame Nullstellen enthält. Dazu betrachten wir die Polynome als univariat in X_n mit Koeffizienten im Ring $\mathcal{R} := \mathbb{Q}[X_1, \ldots, X_{n-1}]$, d.h. $f_1, \ldots, f_s \in \mathcal{R}[X_n]$. Es soll durch Bestimmen von Resultanten die Variable X_n eliminiert werden.

Bei der Bestimmung von parameterabhängigen Resultanten entsteht eine unbestimmte Situation, wenn unter Spezialisierung der Parameter der Grad der beteiligten Polynome absinkt. Um also die Grade der Polynome $f_1, \ldots, f_s \in \mathcal{R}[X_n]$ bei Spezialisierung von $X_1, \ldots, X_{n-1}$ zu einem Tupel komplexer Zahlen zu erhalten, ist es notwendig, dass die führenden Koeffizienten der Polynome in $\mathcal{R}$ invertierbar sind. Das ist der Fall, wenn diese Koeffizienten von Null verschiedenen rationalen Zahlen sind. Davon kann jedoch bei beliebig vorgegebenen Polynomen nicht ausgegangen werden, diese Lage kann aber leicht durch einen Koordinatenwechsel herbeigeführt werden.

Unter einer linearen Transformation der Koordinaten ändert sich lediglich die Lage der Nullstellenmenge, ihre Struktur bleibt erhalten. Wir ersetzen die Variablen $X_1, \ldots, X_n$ auf umkehrbare Weise durch neue Unbestimmte $Y_1, \ldots, Y_n$ nach folgender Vorschrift.

$$\begin{aligned} X_n &:= Y_n, & Y_n &= X_n \\ X_k &:= Y_k + c_k Y_n, & Y_k &= X_k - c_k X_k, & k &= 1, \ldots, n-1. \end{aligned}$$

Für jeden Multiindex $\alpha = (\alpha', \alpha_n) \in \mathbb{N}^n$ wird das Monom $X^\alpha = (X')^{\alpha'} X_n^{\alpha_n}$ unter diesem Variablenwechsel in eine Summe

$$X^\alpha = (Y' + c'Y_n)^{\alpha'} Y_n^{\alpha_n} = \sum_{\substack{\beta', \gamma' \in \mathbb{N}^{n-1} \\ \beta' + \gamma' = \alpha'}} \binom{\alpha'}{\beta'} Y^{\gamma'} c^{(\beta', 0)} \; Y_n^{|\beta'| + \alpha_n}$$

zerlegt, mit den Kurznotationen $X' = (X_1, \ldots, X_{n-1})$, $\alpha' = (\alpha_1, \ldots, \alpha_{n-1})$ und $c = (c', 1)$, $c' = (c_1, \ldots, c_{n-1})$. Weiter ist dabei $\binom{\alpha'}{\beta'} = \binom{\alpha_1}{\beta_1} \ldots \binom{\alpha_{n-1}}{\beta_{n-1}}$.

Wir lesen ab, dass der führende Term in Y_n das Monom $c^\alpha Y_n^{|\alpha|}$ ist. Somit ist für ein Polynom $f_k(X) = \sum_{\alpha \in \mathbb{N}^n} f_{k,\alpha} X^\alpha$ des Grades $\deg f_k$ der führende Koeffizient in Y_n nach Variablentransformation das Monom $Y_n^{\deg f_k} f_k^{(lh)}(c)$, sofern dieser Ausdruck von Null verschieden ist. Dabei bezeichnet $f_k^{(lh)}$ die führende homogene Komponente von f_k,

$$f_k^{(lh)}(X) := \sum_{\alpha \in \mathbb{N}^n:\ |\alpha| = \deg f_k} f_{k,\alpha} X^\alpha \,.$$

Die rationalen Zahlen $f_1^{(lh)}(c), \ldots, f_s^{(lh)}(c)$ sind alle gemeinsam von Null verschieden, wenn deren Produkt

$$h(c) := f_1^{(lh)}(c) \cdots f_s^{(lh)}(c)$$

von Null verschieden ist. Da keiner der Faktoren des Polynoms h identisch zum Nullpolynom ist, definiert das Produkt eine Hyperfläche in $\mathbb{Q}^{n-1}$. Wählt man also ein Tupel $c' = (c_1, \ldots, c_{n-1}) \in \mathbb{Q}^{n-1}$ außerhalb dieser Hyperfläche, so haben alle transformierten Polynome als univariate Polynome in $\mathcal{R}[Y_n]$ einen von Null verschiedenen führenden Koeffizienten, die weiteren Koeffizienten sind aus dem Ring $\mathcal{R} := \mathbb{Q}[Y_1, \ldots, Y_{n-1}]$. Der Einfachheit halber bezeichnen wir das Ergebnis dieser Transformation wieder mit der ursprünglichen Notation, $f_1, \ldots, f_s \in \mathcal{R}[X_n]$, $\mathcal{R} = \mathbb{Q}[X_1, \ldots, X_{n-1}]$ haben nun also führende Koeffizienten in $\mathbb{Q} \setminus \{0\}$.

2.3.2 Berücksichtigung gemeinsamer Faktoren

Gibt es einen gemeinsamen Faktor $r \in \mathcal{R}[X_n]$, so hat dieser auf Grund obiger Bedingung ebenfalls einen führenden rationalen Koeffizienten und kann daher mittels des euklidischen Algorithmus bestimmt werden. Für jede Spezialisierung $\xi \in \mathbb{C}^{n-1}$ der Variablen $X_1, \ldots, X_{n-1}$ erhalten wir ein Polynom $r(\xi)(X_n) \in \mathbb{C}[X_n]$, welches maximal $\deg_{X_n} r$ verschiedene Nullstellen $\eta \in \mathbb{C}$ hat. $(\xi, \eta) \in \mathbb{C}^n$ ist dann eine Lösung des polynomialen Gleichungssystems. Wir erhalten somit einen Teil der Lösungsmenge, der $n-1$ freie Parameter aufweist. Werden diese nur zu rationalen Zahlen spezialisiert, so sind die zugehörigen Werte von X_n algebraisch über $\mathbb{Q}$.

Um den verbleibenden Teil der Lösungsmenge zu bestimmen, dividieren wir alle Polynome durch den gemeinsamen Faktor r. Wir bezeichnen die reduzierten Polynome wieder mit $f_1, \ldots, f_s$, sie sind nun teilerfremd.

2.3.3 Elimination einer Variablen

Um die Variable X_n zu eliminieren und Bedingungen an die verbleibenden Variablen zu erhalten, die die Lösungsmenge vollständig charakterisieren, genügt es nicht, nur die paarweisen Resultanten zu bilden. Es ist vielmehr notwendig, eine genügende Anzahl Paaren von Linearkombinationen der Polynome zu betrachten.

Systematisch wird dies erreicht, indem der Ring $\mathbb{Q}[\underline{X}] = \mathbb{Q}[X_1, \ldots, X_n]$ durch Parameter

$$\underline{U} = (U_1, \ldots, U_s) \quad \text{und} \quad \underline{V} = (V_1, \ldots, V_s)$$

zu $\mathbb{Q}[\underline{X}][\underline{U}, \underline{V}]$ erweitert wird. In $\mathbb{Q}[\underline{X}, \underline{U}, \underline{V}]$ betrachten wir nun die Polynome $g_1 := U_1 f_1 + \cdots + U_s f_s$ und $g_2 := V_1 f_1 + \cdots + V_s f_s$.

Setzt man eine gemeinsame Nullstelle $\xi \in \mathbb{C}^n$ der $f_1, f_2, \ldots, f_s$ in die Polynome g_1, g_2 ein, so werden beide zum Nullpolynom von $\mathbb{C}[\underline{U}, \underline{V}]$. Existiert umgekehrt ein Punkt $\xi \in \mathbb{C}^n$, so dass die Spezialisierungen von $\underline{X}$ in g_1 und g_2 jeweils das Nullpolynom ergeben, so ist dieser Punkt ξ auch eine gemeinsame Nullstelle der Polynome $f_1, f_2, \ldots, f_s$. Ein Punkt $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{C}^n$ ist daher genau dann eine Nullstelle des Systems $f_1, f_2, \ldots, f_s$, wenn die Auswertung der Resultante $\mathrm{Res}_{X_n}(g_1, g_2) \in \mathbb{Q}[X_1, \ldots, X_{n-1}, \underline{U}, \underline{V}]$ in $\xi' := (\xi_1, \ldots, \xi_{n-1})$ das Nullpolynom in $\mathbb{C}[\underline{U}, \underline{V}]$ ergibt.

Wenn wir nun die Resultante nach Monomen in den Parametern $\underline{U}, \underline{V}$ entwickeln, so sind die Koeffizienten dieser Monome Polynome $f'_1, \ldots, f'_{s'} \in \mathbb{Q}[X_1, \ldots, X_{n-1}]$. Das Verschwinden der betrachteten Resultante unter Teilauswertung nach $\underline{X}' = (X_1, \ldots, X_{n-1})$ ist somit äquivalent zum Verschwinden aller dieser Koeffizientenpolynome.

Wir haben nun ein System von Polynomen gewonnen, welches dieselbe Nullstellenmenge charakterisiert, jedoch einen Freiheitsgrad weniger hat. Zu jedem Punkt $\xi' = (\xi_1, \ldots, \xi_{n-1})$ der Nullstellenmenge des Systems $f'_1, \ldots, f'_{s'}$ gibt es einen nichttrivialen gemeinsamen Faktor der Polynome $f_1(\xi', X_n), \ldots, f_1(\xi', X_n) \in \mathbb{C}[X_n]$. Dieser gemeinsame Faktor hat mindestens eine Nullstelle, jede Nullstelle $\xi_n \in \mathbb{C}$ ergibt eine gemeinsame Nullstelle $\xi = (\xi', \xi_n)$ des Systems $f_1, \ldots, f_s$. Umgekehrt ergibt jedes Element $\xi = (\xi', \xi_n) \in \mathbb{C}^n$ der Nullstellenmenge des Systems $f_1, \ldots, f_s$ eine gemeinsame Nullstelle ξ' des Systems $f'_1, \ldots, f'_{s'}$.

Wir können annehmen, dass die Nullstellenmenge des Systems $f'_1, \ldots, f'_{s'}$ nach Komponenten mit $0, 1, \ldots, n-2$ freien Parametern zerlegt ist. Die Anzahl freier Parameter jeder Komponente bleibt beim Übergang zur Nullstellenmenge des Systems $f_1, \ldots, f_s$ erhalten, es kommt zusätzlich noch die Komponente mit $n-1$ freien Parametern des zuvor herausdividierten gemeinsamen Faktors des Ausgangssystems hinzu.

Der soeben skizzierte Eliminationsschritt kann also induktiv zu einem Verfahren zur Elimination aller Variablen fortgesetzt werden. Dabei werden – möglicherweise leere – Komponenten zu $0, 1, \ldots, n-1$ freien Parametern konstruiert. Spezialisiert man in diesen Komponenten die Parameter zu rationalen Zahlen, so ergeben sich Punkte der Nullstellenmenge des Systems, deren Koordinaten algebraisch über $\mathbb{Q}$ sind.

Die Vorgehensweise, wie sie hier angegeben ist, führt zu einem hyperexponenentiellen Wachstums der Anzahl und der Grade der Polynome in den konstruierten Systemen der Eliminationsschritte.

2.4 Gröbner–Basen

Es handelt sich bei *Gröbner–Basen* um ein Werkzeug zum effektiven Rechnen mit Polynomidealen. Eines der Grundprobleme der Algebra, welches mit diesem Werkzeug gelöst werden kann, ist die Frage nach der Zugehörigkeit eines Polynoms f zu einem Ideal $I := \langle g_1, \ldots, g_s \rangle$, aufgespannt von s Polynomen $g_1, \ldots, g_s \in \Bbbk[X_1, \ldots, X_n]$ über einem Körper $\Bbbk$, $n \in \mathbb{N}$. Die effektive Beantwortung dieser Frage ist äquivalent zur effektiven Bestimmung einer Basis des Restklassenrings $\Bbbk[\underline{X}]/_I$. Dies ermöglicht nachfolgend z.B. die Bestimmung der Nullstellenmenge für den Fall, dass diese nulldimensional ist und aus endlich vielen Punkten besteht.

Im Falle einer Variablen ist $\Bbbk[X]$ ein Hauptidealring, d.h. jedes Ideal $\langle g_1, \ldots, g_s \rangle$ wird von einem einzigen Polynom, dem größten gemeinsamen Teiler $ggT(g_1, \ldots, g_s)$ erzeugt,

$$\langle g_1, \ldots, g_s \rangle = \langle ggT(g_1, \ldots, g_s) \rangle \ .$$

Die Zugehörigkeit zu einem Ideal $\langle g \rangle$ kann einfach durch Division mit Rest entschieden werden, indem ein gegebenes $f \in \Bbbk[X]$ in ein Vielfaches von g und einen kleinsten Rest $r := f - q\,g$, $q \in \Bbbk[X]$ zerlegt wird, dabei kann unter allen möglichen Resten der kleinste durch Vergleich der Grade, $\deg r < \deg g$, charakterisiert werden. Eine Basis des Restklassenrings bilden dementsprechend die Monome $1, X, \ldots, X^{\deg g - 1}$.

Im Falle eines Polynomrings in mehreren Variablen scheitert die Definition des kleinsten Rests als erstes daran, dass es keine kanonische Ordnung der Polynome wie im univariaten Fall gibt. So kann es bei der Ordnung nach dem (totalen) Grad mehrere kleinste Reste geben.

Mittels einer „*zulässigen Monomordnung*“ kann dieses Problem überwunden werden, es gibt dann unter den Darstellungen

$$f = q_1\,g_1 + \cdots + q_s\,g_s + r$$

eine eindeutig bestimmte mit kleinstem Rest bzgl. dieser Ordnung. Es verbleibt jedoch das größere Problem, diesen Rest auch bestimmen zu können. Ist das System $(g_1, \ldots, g_s)$ jedoch eine *Gröbner–Basis*, so läßt sich diese Frage wieder auf mehrfache Division mit Rest mit Elementen dieses Systems beantworten. Womit sich letztendlich die Frage der Konstruktion einer Gröbner–Basis zu einem beliebigen vorgegebenen System $(g_1, \ldots, g_s)$ stellt.

Gut verständliche Darstellungen der Grundlagen der Theorie der Gröbner–Basen finden sich in [CLO97] und [GG99], in [BW93] werden die Grundlagen in einem umfassenderen algebraischen Zusammenhang eingeordnet. Für Beweise sei auf diese Werke verwiesen, hier orientieren wir uns an [GG99].

2.4.1 Monomordnungen

Wir erinnern daran, dass eine Halbordnung auf einer Menge M eine zweistellige Relation $\preceq \subset M \times M$ ist, welche

- *reflexiv* ist, d.h. $a \preceq a$ für jedes $a \in M$,
- *transitiv* ist, d.h. $a \preceq b$ und $b \preceq c$ erzwingt $a \preceq c$ für jedes Tripel $a, b, c \in M$, und die
- *antisymmetrisch* ist, d.h. aus $a \preceq b$ und $b \preceq a$ folgt immer $a = b$ für alle Paare $a, b \in M$.

Eine (totale) Ordnung hat die zusätzliche Eigenschaft, dass alle Elemente der Menge zueinander in Beziehung stehen, d.h. für alle $a, b \in M$ eins von $a \preceq b$ oder $b \preceq a$ gilt. Eine Wohlordnung ist eine Ordnung, in welcher jede Teilmenge ein kleinstes Element enthält.

So ist $\leq$ auf $\mathbb{N}$ eine Ordnungsrelation, jedoch die komponentenweise Ausdehnung auf $\mathbb{N}^n$ ist nur eine Halbordnung, dabei stehen zwei Elemente $a, b \in \mathbb{N}^n$ in Beziehung zueinander, $a \leq b$, wenn es ein $c \in \mathbb{N}^n$ mit $a + c = b$ gibt.

Definition 2.4.1 *Für jede Teilmenge $A \subset \mathbb{N}^n$ heiße $A + \mathbb{N}^n := \{a + c : \ a \in A \,\&\, c \in \mathbb{N}^n\}$ die* Stabilisierung *von A. Eine Teilmenge A heiße* stabil*, wenn sie mit ihrer Stabilisierung übereinstimmt.*

Die Menge $A + \mathbb{N}^n$ kann auch als $\{b \in \mathbb{N}^n | \ \exists a \in A : a \leq b\}$ charakterisiert werden. Insbesondere ist jede Menge $A + \mathbb{N}^n$ stabil.

Lemma 2.4.2 (Dicksons Lemma, s. [GG99], Thm. 21.18) *Zu jeder Menge $A \subset \mathbb{N}^n$ gibt es eine endliche Teilmenge $A_0 \subset A$, so dass die Stabilisierungen beider Mengen übereinstimmen, $A_0 + \mathbb{N}^n = A + \mathbb{N}^n$.*

Definition 2.4.3 *Eine Halbordnung $\preceq$ auf der Menge $\mathbb{N}^n$ heiße linear, wenn für $\alpha, \beta, \gamma \in \mathbb{N}^n$ die Beziehung $\alpha \preceq \beta$ äquivalent zu $\alpha + \gamma \preceq \beta + \gamma$ ist.*

Eine Ordnung $\preceq$ auf $\mathbb{N}^n$ heiße zulässig *(im Sinne der Theorie der Gröbner–Basen), wenn sie eine lineare Wohlordnung ist.*

Jede lineare Ordnung ist mit der natürlichen Halbordnung auf $\mathbb{N}^n$ verträglich, d.h. aus $a \leq b$ folgt $a \preceq b$. In einer zulässigen Ordnung ist $0 \in \mathbb{N}^n$ das kleinste Element. Umgekehrt ergibt sich aus dem Lemma von Dickson, dass jede lineare Ordnung auf $\mathbb{N}^n$, für welche 0 das kleinste Element ist, ebenfalls wohlgeordnet ist.

Unter den verschiedenen Möglichkeiten, eine zulässige Ordnung zu definieren, gibt es einige oft gebrauchte. Aufgrund der Linearität solcher Ordnungen muss, um $a \preceq b$ zu entscheiden, nur deren Differenz $d := b - a \in \mathbb{Z}^n$ betrachtet werden.

Definition 2.4.4 *Seien $a, b \in \mathbb{N}^n$, $d := b - a \in \mathbb{Z}^n$. Dann ist $a \prec b$, wenn für die nachfolgend angegebenen Ordnungen die, links beginnend, erste von Null verschiedene Zahl im zugeordneten Tupel positiv ist.*

lexikographisch	*lex*	$d_1, d_2, \ldots, d_n$
graduiert lexikographisch	*grlex*	$\sum_{k=1}^{n} d_k, d_1, d_2, \ldots, d_{n-1}$
revers lexikographisch	*grevlex*	$\sum_{k=1}^{n} d_k, -d_n, -d_{n-1}, \ldots, -d_2$

In der *lexikographischen* Ordnung ist jedes Vielfache des kanonischen Basisvektors e_i kleiner als e_{i+1}, $i = 1, \ldots, n-1$, in einer *graduierten* Ordnung hat dagegen jedes $a \in \mathbb{N}^n$ nur endlich viele Vorgänger.

Da eine zulässige Ordnung eine eindeutige Reihenfolge unter den Monomen in einem jeden Polynom festlegt, wird sie auch *Monomordnung* genannt. Sei eine zulässige Monomordnung $\preceq$ fixiert. Jedes Polynom ist eine Linearkombination von endlich vielen Monomen, unter diesen gibt es ein größtes, d.h. jedes $f \in \Bbbk[X]$ hat eine eindeutige Darstellung

$$f(X) = f_\alpha X^\alpha + \sum_{\beta \in \mathbb{N}^n:\, \beta \prec \alpha} f_\beta X^\beta$$

mit nur endlich vielen von Null verschiedenen Koeffizienten. Dann bezeichnen wir

- mit $\mathrm{LE}(f) := \max_{\preceq}\{\alpha \in \mathbb{N}^n :\ f_\alpha \neq 0\}$ den führenden (engl. „leading“) Exponenten von f,
- mit $\mathrm{LC}(f) := f_{\mathrm{LE}(f)}$ den führenden Koeffizienten,
- mit $\mathrm{LT}(f) := X^{\mathrm{LE}(f)}$ den führenden Term von f und
- mit $\mathrm{LM}(f) := \mathrm{LC}(f)\,\mathrm{LT}(f)$ das führende Monom von f.

Werden diese Abbildungen auf eine Menge G von Polynomen angewendet, insbesondere auf Polynomideale, so werde immer das Nullpolynom aus der Bestimmung der Bildmenge ausgeschlossen, unter $\mathrm{LE}(G)$ ist also genauer $\mathrm{LE}(G \setminus \{0\})$ zu verstehen usw.

Definition 2.4.5 *Sei $G \subset \Bbbk[X]$ eine fixierte Menge von Polynomen und $f \in \Bbbk[X]$ beliebig. Für jede Wahl von $g_1, \ldots, g_s \in G$ und $q_1, \ldots, q_s \in \Bbbk[X]$ mit $\mathrm{LE}(q_k g_k) \preceq \mathrm{LE}(f)$ für alle $k = 1, \ldots, s$ heißt*

$$r := f - q_1 g_1 - \cdots - q_s g_s \in \Bbbk[X]\,,$$

eine Reduktion *von f bzgl. G.*

r heißt Rest *bzw.* vollständige Reduktion *von f bzgl. G, wenn kein führender Term aus $\mathrm{LT}(G)$ ein Monom von r teilt.*

Es kann verschiedene Reste von f bzgl. G geben. Es gibt jedoch Polynommengen G, in denen dieser Rest eindeutig ist.

Lemma 2.4.6 (s. [GG99], Lemma 21.27) *Seien $G \subset \Bbbk[\underline{X}]$ eine endliche Menge von Polynomen und $I := \langle G \rangle$ das von G erzeugte Ideal. Ist die Menge der führenden Exponenten $\mathrm{LE}(I) \subset \mathbb{N}^n$ identisch zur Stabilisierung von $\mathrm{LE}(G)$ im Sinne des Lemmas von Dickson, so ist der Rest unter Reduktion bzgl. G eindeutig.*

Definition 2.4.7 *Eine erzeugende Teilmenge G eines Ideals $I = \langle G \rangle$ heißt* Standardbasis[2] *bzw.* Gröbner-Basis[3] *von I, falls $\mathrm{LE}(I)$ von $\mathrm{LE}(G)$ aufgespannt wird. Der Rest nach vollständiger Reduktion bzgl. G werde mit f rem G bezeichnet.*

Zur – abstrakten – Konstruktion einer Gröbner–Basis genügt es also, von der Menge aller führenden Exponenten eines Ideals eine endliche, erzeugende Teilmenge nach dem Lemma von Dickson zu bestimmen, und dann zu jedem der endlich vielen Exponenten ein Polynom im Ideal zu wählen, in welchem dieser Exponent der führende ist.

Satz 2.4.8 (Hilbertscher Basissatz, s. [GG99], Lemma 21.22 & Thm. 21.23) *Ist $I \subset \Bbbk[\underline{X}]$ ein Ideal, so gibt es eine endliche, das Ideal erzeugende Teilmenge $G \subset I$. Jede Gröbner–Basis erfüllt diese Bedingung.*

Das eigentliche Berechnungsproblem besteht nun darin, dass es keine effektive Möglichkeit gibt, alle Elemente eines Ideals zu durchlaufen, um die kleinsten führenden Exponenten zu bestimmen. Es müssen also gezielt Polynome des Ideals konstruiert werden, deren führende Exponenten Kandidaten für die Menge der erzeugenden Exponenten sind. Dazu ist es ausreichend, Kombinationen von Paaren eines gegebenen Erzeugendensystem zu betrachten, deren jeweils führende Monome sich gegenseitig aufheben.

Als S–Polynom zweier Polynome $g_1, g_2 \subset I$ eines Ideals wird diejenige Linearkombination

$$S(g_1, g_2) := m_2 g_1 - m_1 g_2 \,,$$

mit Monomen $m_1, m_2 \in \Bbbk[\underline{X}]$ bezeichnet, für welche $m_1 \ \mathrm{LM}(g_2) = m_2 \ \mathrm{LM}(g_1)$ gilt und m_1, m_2 minimal in dieser Eigenschaft sind.

Satz 2.4.9 ([GG99],Thm. 21.31) *Eine endliche erzeugende Teilmenge G eines Ideals $I \subset \Bbbk[\underline{X}]$ ist genau dann eine Gröbner–Basis, wenn sich die S–Polynome $S(g_1, g_2)$ für alle $g_1, g_2 \in G$ bzgl. G zu Null reduzieren lassen.*

Satz 2.4.10 (Buchberger–Algorithmus, s. [GG99], Thm. 21.34) *Sei G_0 eine erzeugende Teilmenge eines Ideals I. Wir konstruieren iterativ eine Folge von erzeugenden Teilmengen $\{G_k\}_{k\in\mathbb{N}}$. Sei dazu D_k eine Menge, die für jedes Paar $g_1, g_2 \subset G_k$ einen Rest von $S(g_1, g_2)$ unter Reduktion bzgl. G_k enthält, und sei $G_{k+1} := G_k \cup D_k$.*

Dann wird diese Folge von Polynommengen stationär, d.h. es gibt ein $m \in \mathbb{N}$ mit $G_{m+p} = G_m$ für alle $p \in \mathbb{N}$ und G_m ist eine Gröbner–Basis.

Definition 2.4.11 *Eine Gröbner–Basis G eines Ideals I heißt*

- minimal, *wenn* $\mathrm{LC}(g) = 1$ *und* $\mathrm{LM}(g) \notin \langle \mathrm{LM}(G \setminus \{g\}) \rangle$ *für jedes $g \in G$, und*
- reduziert, *wenn* $g \,\mathrm{rem}(G \setminus \{g\}) = g$ *für alle $g \in G$ gilt.*

Es läßt sich zeigen, dass es zu jeder Monomordnung genau eine reduzierte minimale Gröbner–Basis gibt.

[2] nach H. Hironaka [Hir64] 1964, eingeführt im Kontext der Untersuchung algebraischer Singularitäten
[3] nach B. Buchberger [Buc65] 1965, eingeführt im Kontext der berechenbaren algebraischen Geometrie

Es gibt verschiedene Methoden, um bei der Konstruktion der Gröbner–Basis die Berechnung der Reste in jedem Schritt möglichst effizient zu gestalten. So kann man bestimmten Paaren von Polynomen direkt ansehen, dass der Rest ihres S–Polynoms verschwindet, s. [Buc79, Buc85, GM88]. Auch kann man die Reduktionen der S–Polynome mehrerer Paare aus G_k parallel betrachten, aus diesem Ansatz resultieren die derzeit schnellsten Algorithmen zur Berechnung von Gröbner–Basen von J. C. Faugère, s. [Fau99].

Jedoch ist allen Modifikationen gemeinsam, dass die beste Schranke für den Grad der Elemente einer Gröbner–Basis $d^{O(2^n)}$ beträgt ([MM82], zitiert nach [GG99]). Es gibt Beispiele für Ideale, deren Gröbner–Basen zu beliebigen Monomordnungen mindestens $2^{2^{cn}}$ Elemente aufweist, und unter diesen gibt es Elemente mit Graden größer $2^{2^{cn}}$, für eine Konstante $c > 0$. Für die Laufzeit eines Algorithmus ist der Aufwand zum Aufschreiben des Ergebnisses eine untere Schranke. Es gibt keinen Algorithmus zum Bestimmen einer Gröbner–Basis, für welchen eine obere Schranke für die Laufzeit bekannt ist.

2.4.2 Nullstellenbestimmung mittels Gröbner–Basen

Seien $\Bbbk$ ein Körper der Charakteristik 0 und $\overline{\Bbbk}$ ein algebraischer Abschluss. Neben der Möglichkeit, dass $\Bbbk = \mathbb{Q}$ und $\overline{\Bbbk} \subset \mathbb{C}$ die Menge der algebraischen komplexen Zahlen ist, ist auch häufig der Fall anzutreffen, dass $\Bbbk = \mathbb{Q}(\underline{U})$ der Körper der rationalen Funktionen in Parametern $\underline{U} = (U_1, \ldots, U_m)$ ist.

Definition 2.4.12 *Sei I ein Ideal im Polynomring $\Bbbk[\underline{X}]$ in n Variablen. Wir nennen I nulldimensional, wenn der Restklassenring $\Bbbk[\underline{X}]/I$ eine endliche algebraische Erweiterung von $\Bbbk$ ist.*

In [GRR99] wurde bewiesen, dass folgendes gilt:

- Ist eine Monomordnung und die zugehörige Gröbner–Basis $G \subset I$ gegeben, so ist I nulldimensional genau dann, wenn eine der folgenden Bedingungen erfüllt ist:
  - $\Bbbk[\underline{X}]/I$ ist ein endlichdimensionaler $\Bbbk$–Vektorraum,
  - $A := \mathbb{N}^n \setminus \mathrm{LE}(I)$ ist eine endliche Teilmenge,
  - für jedes $k = 1, \ldots, n$ gibt es ein m mit $X_k^m \in \mathrm{LT}(G)$, d.h. $m\mathbf{e}_k \in \mathrm{LE}(G)$.
- Ist I nulldimensional, so ist die Anzahl der Nullstellen von I in $\overline{\Bbbk}^n$ durch die Anzahl D der Punkte in A beschränkt.
- Weiterhin bilden in diesem Falle die Restklassen zu den Elementen aus $X^A := \{X^a : a \in A\}$ eine Basis von $\Bbbk[\underline{X}]/I$, insbesondere hat der Restklassenring die Dimension D.

Die Beziehung zwischen dem von X^A aufgespannten Vektorraum und der Nullstellenmenge $V(I) \subset \overline{\Bbbk}^n$ ist die folgende. Zu jedem Polynom $f \in \Bbbk[\underline{X}]$ kann eine lineare Abbildung des Restklassenrings von I in sich definiert werden:

$$m_f : \Bbbk[\underline{X}]/I \to \Bbbk[\underline{X}]/I, \quad (h + I) \mapsto (fh + I) .$$

Bezüglich der Basis $X^A = (X^{\alpha_1}, \ldots, X^{\alpha_D})$, $\alpha_1 \prec \alpha_2 \prec \cdots \prec \alpha_D$, hat m_f eine Matrixdarstellung $M_f \in \Bbbk^{D \times D}$, d.h. $m_f(X^{\alpha_k} + I) = \sum_{l=1}^{D} (M_f)_{k,l} (X^{\alpha_l} + I)$.

Die Multiplikationsabbildungen kommutieren miteinander, denn es gilt $m_f m_g = m_{fg} = m_g m_f$. Damit sind aber die Matrizen M_f, $f \in \Bbbk[\underline{X}]$, simultan über $\overline{\Bbbk}$ trigonalisierbar, d.h. es gibt eine Koordinatentransformation $U \in \overline{\Bbbk}^{D\times D}$, bezüglich welcher alle Matrizen $U^{-1}M_f U$ in Jordan–Normalform mit gleicher Struktur der Jordankästchen sind.

Insbesondere gibt es Vektoren $e_k \in \overline{\Bbbk}^D$, $k = 1,\ldots,\delta$, $\delta \leq D$, die für jede der Matrizen M_f Eigenvektoren sind. Die zugehörige Eigenwertabbildung $\lambda_k : \Bbbk[\underline{X}] \to \overline{\Bbbk}$, definiert durch $\lambda_k(f)e_k := M_f e_k$, ist ein Ringhomomorphismus, d.h. eine Auswertungsabbildung zum Punkt $\xi_k := (\lambda_k(X_1),\ldots,\lambda_k(X_n)) \in \overline{\Bbbk}^D$. Für Polynome $h \in I$ aus dem Ideal gilt $m_h = 0$, also insbesondere $h(\xi_k) = 0$, die Punkte ξ_k sind also Nullstellen des Ideals. Umgekehrt kann jeder Nullstelle $x \in V(I)$ der Eigenvektor $(x^{\alpha_1},\ldots,x^{\alpha_D})^t$ zugeordnet werden. Zusammengefasst gilt damit folgender Satz:

Satz 2.4.13 (Stickelbergers Theorem, nach [GRR99, Rou98]) *Ist I ein nulldimensionales Ideal, dann sind die Eigenwerte des Endomorphismus $m_f : \Bbbk[\underline{X}]/I \to \Bbbk[\underline{X}]/I$ die Werte von $f \in \Bbbk[\underline{X}]$ auf den Nullstellen $V(I) \subset \overline{\Bbbk}^n$. Genauer gilt, dass das charakteristische Polynom $\chi_f \in \Bbbk[T]$ von m_f die Linearfaktorzerlegung*

$$\chi_f(T) := \det(TM_1 - M_f) = \prod_{\xi\in V(I)} \big(T - f(\xi)\big)^{\mu(\xi)}$$

hat, wobei $\mu(\xi)$ die Dimension des der Nullstelle ξ zugeordneten Jordan–Kästchens bezeichne.

Seien $f, g_1,\ldots,g_N \in \Bbbk[\underline{X}]$ Polynome und $\underline{Y} := (Y_1,\ldots,Y_N)$ zusätzliche Parameter. Mit diesen seien das Polynom $F(\underline{T},\underline{X}) := f(\underline{X}) + Y_1 g_1(\underline{X}) + \cdots + Y_N g_N(\underline{X}) \in \Bbbk[\underline{Y},\underline{X}]$ und das charakteristische Polynom $Q(\underline{Y},T) := \chi_{F(\underline{Y})}(T)$ für den Multiplikationsoperator $m_{F(\underline{Y})}$ gebildet. Dann gilt in $\overline{\Bbbk}[\underline{Y},T]$

$$\begin{aligned} Q(\underline{Y},T) &:= \det(TM_1 - M_f - Y_1 M_{g_1} - \cdots - Y_N M_{g_N}) \\ &= \prod_{\xi\in V(I)} \big(T - f(\xi) - Y_1 g_1(\xi) - \cdots - Y_N g_N(\xi)\big)^{\mu(\xi)} \end{aligned}$$

Betrachten wir die partielle Ableitungen von Q nach T und $Y_1,\ldots,Y_N$ und setzen dabei $\underline{Y} = 0$, so erhalten wir aus der Produktdarstellung von Q die Identitäten

$$\begin{aligned} \frac{\partial Q(\underline{Y},T)}{\partial T}\Big|_{\underline{Y}=0} &= \chi_f'(T) = \sum_{\xi\in V(I)} \mu(\xi)\big(T - f(\xi)\big)^{\mu(\xi)-1} \prod_{\zeta\in V(I)\setminus\{\xi\}} \big(T - f(\zeta)\big)^{\mu(\zeta)} \\ \frac{\partial Q(\underline{Y},T)}{\partial Y_k}\Big|_{\underline{Y}=0} &= -\sum_{\xi\in V(I)} \mu(\xi) g_k(\xi)\big(T - f(\xi)\big)^{\mu(\xi)-1} \prod_{\zeta\in V(I)\setminus\{\xi\}} \big(T - f(\zeta)\big)^{\mu(\zeta)}. \end{aligned}$$

Allen diesen Polynomen gemeinsam ist der Faktor $\prod_{\xi\in V(I)}\big(T - f(\xi)\big)^{\mu(\xi)-1}$, der sich als größter gemeinsamer Teiler von χ_f und χ_f' in $\Bbbk[\underline{\lambda}][T]$ bestimmen läßt, sofern die Werte von f auf den Nullstellen paarweise verschieden sind.

Definition 2.4.14 *Sei $I \subset \Bbbk[\underline{X}]$ ein nulldimensionales Ideal und $V(I)$ dessen Nullstellenmenge. Ein Polynom $f \in \Bbbk[\underline{X}]$ heißt* separierend, *wenn für $\xi,\eta \in V(I)$ aus $f(\xi) = f(\eta)$ schon $\xi = \eta$ folgt.*

Fast alle Polynome f sind separierend, es ist sogar ausreichend, f linear zu wählen. Denn im Raum der linearen Polynome definieren je zwei Punkte in $\mathbb{k}^n$ eine Hyperebene von Polynomen, die diesen denselben Wert zuordnen. Ist $\mathbb{k}$ von Charakteristik 0, wie vorausgesetzt, so ist das Komplement zu den Hyperebenen zu Paaren aus D Punkten nichtleer.

Sei nun f separierend, so dass $ggT(\chi_f, \chi_f')$ in der Tat der gemeinsame Faktor der oben angegebenen Polynome ist. Dann können wir alle diese Polynome durch $ggT(\chi_f, \chi_f')$ teilen und erhalten Polynome $q, w, v_1, \ldots, v_N \in \mathbb{k}[T]$ mit

$$\begin{aligned} q(T) &:= \tfrac{1}{ggT(\chi_f,\chi_f')}\chi_f(T) = \prod_{\xi\in V(I)} \big(T - f(\zeta)\big), \\ w(T) &:= \tfrac{1}{ggT(\chi_f,\chi_f')}\tfrac{\partial Q(\underline{Y},T)}{\partial T}\big|_{\underline{Y}=0} = \sum_{\xi\in V(I)} \mu(\xi) \prod_{\zeta\in V(I)\setminus\{\xi\}} \big(T - f(\zeta)\big), \\ v_k(T) &:= -\tfrac{1}{ggT(\chi_f,\chi_f')}\tfrac{\partial Q(\underline{Y},T)}{\partial Y_k}\big|_{\underline{Y}=0} = \sum_{\xi\in V(I)} \mu(\xi) g_k(\xi) \prod_{\zeta\in V(I)\setminus\{\xi\}} \big(T - f(\zeta)\big)\,. \end{aligned}$$

Da die Werte $f(\xi)$, $\xi \in V(I)$ paarweise verschieden sind, liest man aus den obigen Gleichungen ab, dass jeder Funktionswert $\alpha := f(\xi)$ von f in einer Nullstellen von I eine Nullstelle von q ist, d.h. algebraisch über $\mathbb{k}$ ist, und die Werte der anderen Funktionen sich durch rationale Funktionen in α ausdrücken lassen, d.h. sich als $g_k(\xi) = \frac{v_k(\alpha)}{w(\alpha)}$ ergeben. Dies gewinnt Interesse dadurch, dass sich die benutzten Polynome direkt berechnen lassen, ohne die Nullstellen zu kennen.

Insbesondere interessant ist der Fall $N = n$ und $g_k(X) := X_k$. Es ist zwar möglich, die Koordinaten der Nullstellen von I als Nullstellen der charakteristischen Polynome χ_{X_k} zu gewinnen, allerdings ist dann deren Zuordnung zueinander unbekannt, aus δ^n möglichen Kombinationen müssten die δ zutreffenden ausgewählt werden. Mit obiger Konstruktion jedoch erhalten wir die Nullstellen als Brüche $\xi = \left(\frac{v_1(\alpha)}{w(\alpha)}, \ldots, \frac{v_n(\alpha)}{w(\alpha)}\right)$, wobei α eine Lösung von $q(\alpha) = 0$ ist.

2.5 Die TERA–Methode

Wir betrachten wieder das Problem, zu n Polynomen $f_1, \ldots, f_n \in \mathbb{Q}[X_1, \ldots, X_n]$ in n Variablen diejenigen Punkte $x \in \mathbb{C}^n$ zu finden, welche $f_1(x) = \cdots = f_n(x) = 0$ erfüllen. Das hier darzustellende Verfahren gestattet es, auch Ungleichungen der Art $g(x) \neq 0$ im System direkt, d.h. ohne Hinzufügen von Hilfsvariablen, zu berücksichtigen. Mehrere Ungleichungen können mittels Produktbildung zu einer Ungleichung zusammengefasst werden. Es seien also $f_1, \ldots, f_n, g \in \mathbb{Q}[X_1, \ldots, X_n]$ gegeben und aus diesen das System $\mathcal{S}$:

$$x \in \mathbb{C}^n : \quad f_1(x) = \cdots = f_n(x) = 0, \quad g(x) \neq 0$$

gebildet. Diesem System kann die algebraische Varietät V zugeordnet werden, die der algebraische Abschluss der Menge $V(f_1, \ldots, f_n) \setminus V(g)$ ist, d.h. aus allen irreduziblen Komponenten von $V(f_1, \ldots, f_n)$ besteht, die nicht vollständig in der Hyperfläche $V(g)$ enthalten sind.

Der Ansatz der innerhalb der TERA–Gruppe entwickelten effizienten Variante der Kronecker–Methode besteht darin, mit einem, bis auf die Ungleichung $g(x) \neq 0$, leeren System $\mathcal{S}_0$ zu starten und schrittweise die Anzahl der Freiheitsgrade durch Hinzufügen einer Gleichung zum

Gleichungssystem zu reduzieren. In jedem Schritt k, $k = 0, \ldots, n$ wird somit ein Gleichungssystem $\mathcal{S}_k$:

$$x \in \mathbb{C}^n : \quad f_1(x) = \cdots = f_s(x) = 0, \quad g(x) \neq 0$$

betrachtet, zu dessen algebraischer Varietät

$$\mathcal{V}_k := \overline{V(f_1, \ldots, f_k) \setminus V(g)} \qquad \text{(alg. Abschluss)}$$

es eine Parametrisierung à la Kronecker der Dimension $r := n - k$ geben soll. Dieser Zusammenhang zwischen Anzahl der Gleichungen und Dimension der Zwischenvarietäten $\mathcal{V}_1, \ldots, \mathcal{V}_n$ ist garantiert, wenn die Polynome $f_1, \ldots, f_n$ eine *reguläre reduzierte Folge* bilden. Diese Einschränkung kann durch sorgfältige Analyse der äquidimensionalen Komponenten der Varietäten und deren algebraischen Vielfachheiten wieder aufgehoben werden (s. [Lec01b]), soll aber für die hier dargelegte Skizze des Verfahrens vorausgesetzt werden.

Eine weitere Methode, die Komplexität des Lösungsverfahrens unter Kontrolle zu halten, liegt in der Beobachtung begründet, dass man die Parametrisierung einer algebraischen Varietät V in Noether–Normalform einer Dimension r auf die Betrachtung einer generischen Faser einschränken kann. Spezialisiert man die r freien Variablen der Parametrisierung zu Konstanten, so erhält man eine nulldimensionale algebraische Varietät, die eine Faser der Projektion der algebraischen Varietät auf die ersten r Koordinaten ist. Vermeidet das r–Tupel der Konstanten eine bestimmte Hyperfläche, so kann mit dem Newton–Hensel–Verfahren aus der Spezialisierung der Parametrisierung auf die Faser die Parametrisierung zurückgewonnen werden.

Der Ablauf der TERA–Kronecker–Methode kann also wie folgt beschrieben werden:

- Es werden Polynome $a_1, \ldots, a_n \in \mathbb{Q}[X_1, \ldots, X_n]$ vom Grad 1 gewählt, so dass das lineare Gleichungssystem $a_1(x) = \cdots = a_n(x) = 0$ genau eine Lösung besitzt. Die Folge $a_1, \ldots, a_n$ sei weiter in allgemeiner Lage zur Folge $f_1, \ldots, f_n$ des zu lösenden polynomialen Gleichungssystems außerhalb $V(g)$.

- Für jedes $k = 0, \ldots, n$ betrachtet man das System $\mathcal{S}'_k$:

  $$x \in \mathbb{C}^n : \quad f_1(x) = \cdots = f_k(x) = a_{k+1}(x) = \cdots = a_n(x) = 0, \quad g(x) \neq 0$$

  Das System $\mathcal{S}'_0$ enthält das lineare Gleichungssystem $a_1(x) = \cdots = a_n(x) = 0$, welches einfach gelöst werden kann. Eine Teilforderung an die „allgemeine Lage“ der $a_1, \ldots, a_n$ ist, dass die Lösungsmengen der Systeme $\mathcal{S}'_k$ nulldimensional seien und eine maximale Anzahl von Punkten enthalten. Insbesondere soll der Lösungspunkt des linearen Teils von $\mathcal{S}'_0$ nicht in $V(g)$ enthalten sein.

- Um von der Lösungsmenge von $\mathcal{S}'_k$ zu der von $\mathcal{S}'_{k+1}$ zu gelangen, wird die eindimensionale Lösungsmenge des Systems $\mathcal{S}''_k$:

  $$x \in \mathbb{C}^n : \quad f_1(x) = \cdots = f_k(x) a_{k+2}(x) = \cdots = a_n(x) = 0, \quad g(x) \neq 0$$

  konstruiert. Dies geschieht mittels des Newton–Hensel–Verfahrens, einer Kombination des Newton–Verfahrens mit dem Hensel–Lifting. Eine weitere Forderung an die „allge-

meine Lage“ der $a_1, \ldots, a_n$ ist, dass der Newtonschritt ausgeführt werden kann. Insbesondere seien $Y_1 := a_n(X), \ldots, Y_{n-k} := a_{n-k+1}(X)$ die freien Variablen einer Noether–Normalisierung der Dimension $r = n - k$ von $\mathcal{V}_k$.

- Die so erhaltenen Kurven werden mit der Hyperfläche $V(f_{k+1})$ geschnitten. Durch entfernen von in $V(g)$ liegenden Punkten entsteht die Lösungsmenge von $\mathcal{S}'_{k+1}$. Eine weitere Forderung an die „allgemeine Lage“ der $a_1, \ldots, a_n$ ist, dass dabei nur Punkte entfernt werden, die nicht in $\mathcal{V}_k$ liegen.

2.5.1 Simultane Noether–Normalisierung

Seien $\mathbb{k}$ ein Körper der Charakteristik 0, $\overline{\mathbb{k}}$ ein algebraischer Abschluss von $\mathbb{k}$ und $f_1, \ldots, f_n, g \in \mathbb{k}[\underline{X}] = \mathbb{k}[X_1, \ldots, X_n]$, so dass die Varietäten

$$\mathcal{V}_k := \overline{V(f_1, \ldots, f_k) \setminus V(g)}, \quad k = 1, \ldots, n$$

eine Noether–Normalisierung der Dimension $n - k$ besitzen. Analog zur originalen Kronecker–Methode kann als Matrix der Noether-Normalisierung jeder der algebraischen Varietäten $\mathcal{V}_1, \ldots, \mathcal{V}_n$ eine obere Dreiecksmatrix mit Werten 1 auf der Diagonale gewählt werden. Dabei müssen die Einträge der Matrix die Nullstellenmengen einer gewissen Anzahl von Polynomen vermeiden.

Man kann nun einen Schritt weiter gehen und die Bedingungen an die Matrizen der Noether-Normalisierung aller Zwischenvarietäten $\mathcal{V}_1, \ldots, \mathcal{V}_n$ gemeinsam betrachten. Dies entspricht der Vermeidung einer Hyperfläche, die zum Produkt aller beteiligten Polynome gehört. Ist eine solche *simultane Noether–Normalisierung* A gefunden, so befinden sich die transformierten Varietäten $\mathcal{V}_k^A = A^{-1}\mathcal{V}_k$ in Noether–Normalform der Dimension $r = n - k$, $k = 1, \ldots, n$.

In der Definition der Noether–Normalisierung wird die Koordinatentransformation $(X_1, \ldots, X_n)^t = A(Y_1, \ldots, Y_n)^t$ betrachtet. Für jedes $k = 1, \ldots, n$ und $r := n - k$, sind die Variablen $Y_1, \ldots, Y_r$ frei im Koordinatenring der Varietät $\mathcal{V}_k^A$. Die Bedingung, dass Y_{r+1} ein *primitives Element* der Noether–Normalisierung von $\mathcal{V}_k^A$ sei, ist wieder durch das Nichtverschwinden eines Polynoms in den Einträgen der Matrix A ausdrückbar. Es können also auch diese Bedingungen zu den schon vorhandenen hinzugefügt werden, wodurch sich die zu vermeidende Hyperfläche um weitere Komponenten vergrößert. Ist der Grundring des Polynomrings wie hier ein Körper $\mathbb{k}$ der Charakteristik 0, z.B. $\mathbb{k} = \mathbb{Q}$, so ist das Komplement einer Hyperfläche nie leer. Es gibt also immer Matrizen, die die oben aufgestellten Bedingungen erfüllen. Diese Matrizen nennt man auch *generisch* (d.h. in einer offenen dichten Teilmenge bzgl. einer geeignet gewählten Zariski–Topologie liegend).

Es kann also im Folgenden immer davon ausgegangen werden, dass – nach einer generisch gewählten Koordinatentransformation – die Zwischenvarietäten $\mathcal{V}_k = V(f_1, \ldots, f_k)$ sich in Noether–Normalform der Dimension $r = n - k$ mit freien Variablen $X_1, \ldots, X_r$ und primitivem Element X_{r+1} befinden. Es gibt somit Polynome $q_k, v_{k,r+2}, \ldots, v_{k,n} \in \mathbb{k}[X_1, \ldots, X_r][Y]$, so

dass jeder Punkt $x = (x_1, \ldots, x_n) \in \mathcal{V}_k$ das Gleichungssystem

$$\begin{aligned} 0 =& q(x_1, \ldots, x_r)(x_{r+1}) \\ 0 =& \frac{\partial q}{\partial Y}(x_1, \ldots, x_r)(x_{r+1})\, x_{r+2} - v_{k,r+2}(x_1, \ldots, x_r)(x_{r+1}) \\ & \vdots \\ 0 =& \frac{\partial q}{\partial Y}(x_1, \ldots, x_r)(x_{r+1})\, x_n - v_{k,n}(x_1, \ldots, x_r)(x_{r+1}) \end{aligned}$$

erfüllt und bei $\frac{\partial q}{\partial Y}(x_1, \ldots, x_r)(x_{r+1}) \neq 0$ auch aus diesem bestimmt werden kann.

2.5.2 Lifting–Faser und Newton–Hensel–Verfahren

Eine Methode der angewandten Mathematik zur Lösung nichtlinearer Gleichungssysteme sind die Homotopieverfahren. Betrachten wir z.B. folgende Situation. Es sei ein Gleichungssystem aus stetig differenzierbaren Funktionen $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ gegeben. Auf dessen Lösungsmenge $M = \{x \in \mathbb{R}^n : f_1(x) = \cdots = f_k(x) = 0\}$ werden die Nullstellen einer weiteren, stetig differenzierbaren Funktion $f_{k+1} : \mathbb{R}^n \to \mathbb{R}$ gesucht. Es sei schon eine Lösung $x^0 \in \mathbb{R}^n$ des Gleichungssystems bekannt, mit diesem Punkt wird eine Hilfsfunktion

$$h_t(x) := f_{k+1}(x) - (1-t)f(x^0)$$

konstruiert. Der Punkt $x^0 \in M$ ist Lösung von $h_0(x) = 0$. Durch aufeinanderfolgende kleine Erhöhungen von t und Korrektur des Lösungspunktes versucht man, für $0 < t_1 < t_2 < \cdots < t_N = 1$ Punkte $x^k \in M$ zu finden, die Lösungen von $h_{t_k}(x^k) = 0$ sind, x^N ist somit einer der gesuchten Lösungspunkte mit $f_1(x^N) = \cdots = f_k(x^N) = f_{k+1}(x^N) = 0$.

Wichtig für jedes Homotopieverfahren ist, dass das Gleichungssystem lokal invertierbar ist, d.h. dass die Punkte auf dem Pfad der Zwischenlösungen *regulär* sind.

Definition 2.5.1 *Seien $\mathbb{k}$ ein Körper, $\overline{\mathbb{k}}$ ein algebraischer Abschluss von $\mathbb{k}$ und $f_1, \ldots, f_k \in \mathbb{k}[X_1, \ldots, X_n]$ Polynome. Ein Punkt $x \in \overline{\mathbb{k}}^n$ heißt* $(f_1, \ldots, f_k)$–regulär, *wenn er Nullstelle des Gleichungssystems $f_1(x) = \cdots = f_k(x) = 0$ ist und die Jacobi–Matrix*

$$J_f(x) := \frac{\partial(f_1, \ldots, f_k)}{\partial(X_1, \ldots, X_n)}(x) = \begin{pmatrix} \frac{\partial f_1}{\partial X_1}(x) & \cdots & \frac{\partial f_1}{\partial X_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_k}{\partial X_1}(x) & \cdots & \frac{\partial f_k}{\partial X_n}(x) \end{pmatrix} \in \overline{\mathbb{k}}^{k \times n}$$

den vollen Rang k besitzt.

Auf endliche Weise darstellbare Punkte einer algebraischen Varietät treten meist in der Mehrzahl auf. Eine derartige endliche, mehrere Punkte der Varietät enthaltende Darstellung wird durch die Fasern einer Parametrisierung à la Kronecker definiert. Der Begriff des regulären Punktes muss für diese Situation erweitert werden.

Definition 2.5.2 *Seien $\mathbb{k}$ ein Körper der Charakteristik 0 und $\overline{\mathbb{k}}$ ein algebraischer Abschluss von $\mathbb{k}$. Seien weiter $f_1, \ldots, f_k \in \mathbb{k}[X_1, \ldots, X_n]$ Polynome, deren Ideal $I := \langle f_1, \ldots, f_k \rangle$ radikal ist und deren Varietät $\mathcal{V}_k = V(f_1, \ldots, f_k) \subset \overline{\mathbb{k}}^n$ sich in Noether–Normalform der Dimension $r := n - k$ befindet.*

Ein Punkt $p \in \mathbb{k}^r$ heißt Lifting–Punkt*, wenn die Faser über p der Projektion $\pi_r : \mathcal{V}_k \to \overline{\mathbb{k}}^r$, $x = (x_1, \ldots, x_n) \mapsto \pi(x) := (x_1, \ldots, x_r)$ aus $(f_1, \ldots, f_k)$–regulären Punkten besteht. $\pi_r^{-1}(p) \cap \mathcal{V}_k$ wird* Lifting–Faser *genannt.*

Diese Definition kann auch auf die Varietät $\mathcal{V}_k$ eines Systems $f_1(x) = \cdots = f_k(x) = 0$, $g(x) \neq 0$ ausgedehnt werden. In Fasern, in welchen sich $\mathcal{V}_k$ und $V(g)$ schneiden, sind die Schnittpunkte der Varietät mit der Hyperfläche singulär. Daher können diese Fasern keine Lifting–Fasern sein. Alle weiteren Fasern über Punkten $p \in \mathbb{k}^r$ entstehen dadurch, dass von der Faser $\pi_r^{-1}(p) \cap V(f_1, \ldots, f_k)$ die in $V(g)$ liegenden Punkte abgezogen werden.

Aufgrund der Regularitätsforderung kann die Lifting–Faser mittels des Implizite–Funktionen–Theorems auf eine Umgebung des Punktes p fortgesetzt werden. Für eine algebraische Varietät ist das aber schon gleichbedeutend mit der Kenntnis der Varietät außerhalb einer Hyperfläche.

Wird die Noether–Normalisierung von $\mathcal{V}_k$ mit einer passend gewählten Koordinatentransformation vorgenommen, so kann neben der Noether–Normalform der Dimension r ebenfalls vorausgesetzt werden, dass X_{r+1} ein primitives Element ist. Es gibt also Polynome

$$Q, V_{r+2}, \ldots, V_n \in \mathbb{k}[X_1, \ldots, X_r][Y] \,,$$

mit welchen jeder Punkt $x = (x_1, \ldots, x_n) \in \mathcal{V}_k$ das Gleichungssystem

$$\begin{aligned} Q(x_1, \ldots, x_r)(x_{r+1}) &= 0 \\ \frac{\partial Q}{\partial Y}(x_1, \ldots, x_r)(x_{r+1}) x_m &= V_m(x_1, \ldots, x_r)(x_{r+1}), \quad m = r+2, \ldots, n, \end{aligned}$$

erfüllt. Ist $\rho := \mathrm{Disk}(Q)$ die Diskriminante von Q nach Y und gilt $\rho(p_1, \ldots, p_r) \neq 0$ für ein $p \in \overline{\mathbb{k}}^r$, so sind durch dieses Gleichungssystem auch alle Punkte von $\mathcal{V}_k$ eindeutig charakterisiert, deren erste r Koordinaten mit p übereinstimmen.

Sei $p \in \mathbb{k}^r$ ein Punkt, der Liftingpunkt der Noether–Normalform ist und gleichzeitig $\rho(p) \neq 0$ erfüllt. Die Spezialisierung der Parametrisierung $Q, V_{r+2}, \ldots, V_n$ in p ergibt eine geometrische Lösung der Faser über p, d.h. Polynome $q, v_{r+2}, \ldots, v_n \in \mathbb{k}[Y]$, mit welchen die Punkte der Faser über p die Lösungen des Systems

$$\begin{aligned} x_1 = p_1, \ldots, x_r &= p_r, \\ q(x_{r+1}) &= 0, \\ q'(x_{r+1}) x_{r+2} = v_{r+2}(x_{r+1}), \ldots, q'(x_{r+1}) x_n &= v_n(x_{r+1}) \end{aligned}$$

sind.

Seien

$$\mathfrak{m} := \langle\, X_1 - p_1, \ldots, X_r - p_r \,\rangle \subset \mathbb{k}[X_1, \ldots, X_r] \quad \text{und} \quad \mathcal{R}_N := \mathbb{k}[X_1, \ldots, X_r]/\mathfrak{m}^N$$

für jedes $N \in \mathbb{N}$. Dann können $q, v_{r+2}, \ldots, v_n$ auch als gradminimale Repräsentanten der Restklassen von $Q, V_{r+2}, \ldots, V_n$ im Restklassenring $\mathcal{R}_1$ interpretiert werden. Ist N größer als

der Grad von Q, so sind $Q, V_{r+2}, \ldots, V_n$ schon selbst die gradminimalen Repräsentanten der von ihnen definierten Restklassen in $\mathcal{R}_N$.

Stellt man sich nun die Aufgabe, in umgekehrter Richtung von einer Lifting–Faser zur Parametrisierung der Varietät $\mathcal{V}_k$ zu gelangen, so müssen aus den Restklassen in $\mathcal{R}_1$ der Polynome der geometrischen Lösung die zugehörigen Restklassen in $\mathcal{R}_N$ bestimmt werden, mit einem ausreichend großen $N \in \mathbb{N}$. Dies geschieht mittels des Newton–Hensel–Verfahrens. In diesem wird die gegebene geometrische Lösung der Lifting–Faser als Parametrisierung von $\mathcal{V}_k$ in $\mathcal{R}_1[Y]$, d.h. mit Fehler in $\mathfrak{m}$, interpretiert. Die Genauigkeit dieser Parametrisierung wird dann rekursiv verbessert. Dabei wird in jedem Rekursionsschritt aus einer Parametrisierung in $\mathcal{R}_N[Y]$ mit Fehler in $\mathfrak{m}^N$ eine Parametrisierung von $\mathcal{V}_k$ in $\mathcal{R}_{2N}[Y]$ gewonnen.

Betrachten wir den Rekursionsanfang, d.h. die Parametrisierung von $\mathcal{V}_k$ in $\mathcal{R}_1[Y]$ Nach Voraussetzung ist $q'(Y)$ invertierbar modulo $q(Y)$, das inverse Element sei $A \in \Bbbk[Y]$. Seien $w_1, \ldots, w_n \in \Bbbk[Y]$ definiert als

$$w_1 := p_1, \ldots, w_r := p_r, \quad w_{r+1} := Y \quad \text{und} \quad w_{r+2} := A(Y)v_{r+2}(Y), \ldots, w_n := A(Y)v_n(Y)\,.$$

Dann gelten:

- $f_1(w) = \cdots = f_k(w) = 0 \bmod \langle\, q(Y) \,\rangle$ und
- die Jacobi–Matrix $J_f(w)$ ist invertierbar modulo $\langle\, q(Y) \,\rangle$.

Die Polynome $w_1, \ldots, w_r$ gehören in $\mathcal{R}_1$ denselben Restklassen wie die Polynome $X_1, \ldots, X_r$ an. Genauso können auch $q, w_{r+2}, \ldots, w_n$ als Polynome in $\mathcal{R}_1[Y]$ betrachtet werden.

Im weiteren wird ein Rekursionsschritt des Newton–Hensel–Verfahrens dargestellt. Seien $N \in \mathbb{N}$ und $Q, W_{r+2}, \ldots, W_n \in \mathcal{R}_N[Y]$, so dass sie modulo $\mathfrak{m}$ mit $q, w_{r+2}, \ldots, w_n$ übereinstimmen und gleichzeitig

$$f(W) = (f_1(W), \ldots, f_k(W)) = 0 \bmod \langle\, Q \,\rangle$$

gilt. Dabei ist $W = (W_1, \ldots, W_n) \in \mathcal{R}_N[Y]$ und analog zu oben $W_1 = X_1 + \mathfrak{m}^N$, $\ldots$, $W_r = X_r + \mathfrak{m}^N$ sowie $W_{r+1} = Y$. Weiter ist $J_f(W) = \frac{\partial(f_1,\ldots,f_k)}{\partial(x_{x-1},\ldots,x_n)}(W)$ invertierbar modulo $\langle\, Q \,\rangle$, da im gegenteiligen Fall dies auch für $N = 1$ nicht gelten würde.

Es kann somit der Newton–Operator auf das Gleichungssystem angewandt werden. Seien $Q, W_{r+2}, \ldots, W_n$ zunächst in ihre gradminimalen polynomialen Repräsentanten und diese dann in die zugehörigen Restklassen in $\mathcal{R}_{2N}[Y]$ transformiert. Mit diesen sei

$$(\tilde{W}_{r+1}, \ldots, \tilde{W}_n) := (W_{r+1}, \ldots, W_n) - J_f(W)^{-1} f(W) \bmod \langle\, Q \,\rangle\,.$$

Sei wieder $\tilde{W}_1 = X_1 + \mathfrak{m}^{2N}, \ldots, \tilde{W}_r = X_r + \mathfrak{m}^{2N}$. Entwickelt man $f(W + h)$ in das Taylorpolynom bzgl. h und berücksichtigt, dass das Inkrement $h := \tilde{W} - W$ des Newton–Operators in $\mathfrak{m}^N$ liegt, so folgt

$$f(\tilde{W}) = 0 \bmod \langle\, Q \,\rangle\,.$$

Die so gefundene genauere Lösung muss nun auf die Form der Parametrisierung von $\mathcal{V}_k$ korrigiert werden, d.h. X_{r+1} soll wieder das primitive Element sein. Es gilt aber $\tilde{W}_{r+1}(Y) =$

$Y + \Delta(Y)$, nach Konstruktion sind die Koeffizienten von $\Delta \in \mathcal{R}_{2N}[Y]$ in $\mathfrak{m}^N$ enthalten. Somit kann zu einer neuer Variablen T übergegangen werden und $Y := T - \Delta(T) \in \mathcal{R}_{2N}[T]$ als Polynom neu definiert werden. Dann gilt $T = Y(T) + \Delta(Y(T))$, definiert man also

$$\begin{aligned} Q^+(T) :=& Q(T - \Delta(T)) = Q(T) - \left(Q'(T)\Delta(T) \bmod Q(T)\right) \quad \text{und} \\ W^+(T) :=& \tilde{W}(Y(T)) = \tilde{W}(T) - \left(\tilde{W}'(T)\Delta(T) \bmod Q(T)\right), \end{aligned}$$

so gelten $W_1^+ = X_1 + \mathfrak{m}^{2N}, \ldots, W_r^+ = X_r + \mathfrak{m}^{2N}$ und $W_{r+1}^+ = T$ sowie

$$f(W^+(T)) = 0 \bmod \langle\, Q^+(T) \,\rangle\ .$$

Da die vorgenommenen Korrekturen sämtlich der N–ten Potenz $\mathfrak{m}^N$ des Ideals $\mathfrak{m} = \langle\, X_1 - p_1, \ldots, X_r - p_r \,\rangle$ angehören, reduzieren sich auch $Q^+, W_{r+2}^+, \ldots, W_n^+$ modulo $\mathfrak{m}$ zu den Polynomen $q, w_{r+2}, \ldots, w_n$ der geometrischen Lösung der Liftingfaser über p.

Somit sind die Voraussetzungen für einen weiteren Schritt des Newton–Hensel–Verfahrens gegeben, jedoch in $\mathcal{R}_{2N}[Y]$ statt in $\mathcal{R}_N$. Ersetzt man N durch $2N$ und Q, W durch Q^+, W^+, so kann dieser Schritt wiederholt werden.

Nach einer endlichen Anzahl von Wiederholungen dieses Schrittes ist N größer als der Grad von q. Man kann zeigen, dass dann der gradminimale Repräsentant von Q, der weiterhin mit $Q \in \Bbbk[X_1, \ldots, X_r][Y]$ bezeichnet sei, das Minimalpolynom von X_{r+1} in der Noether–Normalform der algebraischen Varietät $\mathcal{V}_k$ ist. Weiter sind die gradminimalen Repräsentanten V_m der Restklasse $Q'(Y)W_m(Y) + \langle\, Q(Y) \,\rangle$ aus $\mathcal{R}_N[Y]/\langle\, Q(Y) \,\rangle$ gerade die Koordinatenpolynome der Parametrisierung à la Kronecker von $\mathcal{V}_k$, für exakte Beweise und weitere Details s. [Leh99, Lec01b, GLS01].

Eine Faser über einem Punkt $p \in \Bbbk^r$ ist bestimmt eine Lifting–Faser, wenn alle Punkte der Faser von $\mathcal{V}_k$ bzgl. des erweiterten nulldimensionalen Systems

$$f_1(x) = \cdots = f_k(x) = x_1 - p_1 = \cdots = x_r - p_r = 0$$

regulär sind. Dies ist äquivalent dazu, dass die Jacobi–Matrizen

$$\frac{\partial(f_1, \ldots, f_k)}{\partial(X_{r+1}, \ldots, X_n)}(x) = \begin{pmatrix} \frac{\partial f_1}{\partial X_{r+1}}(x) & \cdots & \frac{\partial f_1}{\partial X_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_k}{\partial X_{r+1}}(x) & \cdots & \frac{\partial f_k}{\partial X_n}(x) \end{pmatrix} \in \overline{\Bbbk}^{k \times k}$$

in diesen Punkten eine nichtverschwindende Determinante $\Delta(x)$ besitzen. Diese Determinante ist ein Polynom, ihr Nichtverschwinden in allen Punkten der Faser ist äquivalent zum Nichtverschwinden des konstanten Koeffizienten des Minimalpolynoms von Δ bzgl. der Noether-Normalform von $\mathcal{V}_k$. Das Minimalpolynom ist in $\Bbbk[X_1, \ldots, X_r][Y]$ enthalten, der konstante Koeffizient daher ein Polynom in $\Bbbk[X_1, \ldots, X_r]$. Dass dieser Koeffizient nicht identisch zum Nullpolynom ist, folgt nach dem Jacobi–Kriterium aus der Radikalität des Ideals $\langle\, f_1, \ldots, f_k \,\rangle$ und der Existenz einer Noether–Normalform.

Betrachtet man nun einen Punkt $p = (p_1, \ldots, p_n) \in \Bbbk^n$, so ist für jedes $k \in \{1, \ldots, n\}$ und $r = n - k$ die Projektion $\pi_r(p) = (p_1, \ldots, p_r) \in \Bbbk^{n-k}$ auf die ersten r Koordinaten ein Lifting–Punkt für $\mathcal{V}_k$, wenn die Hyperfläche eines Polynoms in den ersten r Variablen vermieden

wird. Das Produkt aller dieser Polynome für $k = 1, \ldots, n$ definiert wieder eine Hyperfläche in $\mathbb{k}^n$. Da $\mathbb{k}$ unendlich viele Elemente enthält, gibt es mindestens einen Punkt außerhalb dieser Hyperfläche.

Damit die Lifting–Punkte $\pi_r(p)$ mit der gewählten Parametrisierung von $\mathcal{V}_k$, $k = r - n$, mit X_{r+1} als primitivem Element verträglich sind, muss die Diskriminante $\rho = \mathrm{Disk}(Q)$ des Minimalpolynoms Q von X_{r+1} in $\pi_r(p)$ von Null verschieden sein. Die Diskriminante ist vom Nullpolynom verschieden. Soll diese Bedingung für alle Varietäten $\mathcal{V}_1, \ldots, \mathcal{V}_n$ erfüllt sein, so vergrößert sich die von p zu vermeidende Hyperfläche um eine durch ein Produkt von Polynomen definierte Hyperfläche. Wieder gibt es generische Punkte $p \in \mathbb{k}^n$ (im Sinne einer Zariski–Topologie) außerhalb dieser vergrößerten Hyperfläche.

Ist ein solcher generischer Punkt gewählt, so kann er in den Nullpunkt verschoben werden. Die Kombination aus der Koordinatentransformation und dieser Verschiebung ist eine affine Abbildung. Für ein System

$$x \in \overline{\mathbb{k}}^n : \quad f_1(x) = \cdots = f_n(x) = 0, g(x) \neq 0$$

mit Polynomen $f_1, \ldots, f_n, g \in \mathbb{k}[X_1, \ldots, X_n]$, dessen Zwischenvarietäten

$$\mathcal{V}_k = \overline{V(f_1, \ldots, f_k) \setminus V(g)}, \quad k = 1, \ldots, n,$$

allesamt Noether–normalisierbar mit Dimension $r = n - k$ sind, kann also eine generische affin–lineare Transformation $x = Ay + p$ gefunden werden, so dass mit $f_k^{(A,p)}(y) := f_k(Ay + p)$ und $g^{(A,p)}(y) := g(Ay + p)$ die Zwischenvarietäten $\mathcal{V}_k^{(A,p)}$ des transformierten Systems

$$y \in \overline{\mathbb{k}}^n : \quad f_1^{(A,p)}(y) = \cdots = f_n^{(A,p)}(y) = 0, g^{(A,p)}(y) \neq 0$$

sich in Noether–Normalform der entsprechenden Dimension $r = n - k$ befinden. Ferner, dass $0 \in \mathbb{k}^r$ ein Lifting–Punkt für $\mathcal{V}_k^{(A,p)}$ ist und X_{r+1} sowohl für die Zwischenvarietät als auch für die Faser dieser Varietät über $0 \in \mathbb{k}^r$ ein primitives Element ist.

2.5.3 Schnitt einer Varietät mit einer Hyperfläche

Seien $\mathbb{k}$ ein Körper der Charakteristik 0, $\overline{\mathbb{k}}$ ein algebraischer Abschluss von $\mathbb{k}$ und $f_1, \ldots, f_n, g \in \mathbb{k}[X_1, \ldots, X_n]$ Polynome, so dass die Ideale $\langle f_1, \ldots, f_k \rangle$, $k = 1, \ldots, n$ radikal sind. Sei weiter angenommen, dass bereits eine generische affin–lineare Transformation der Variablen vorgenommen wurde, so dass für jedes $k = 1, \ldots, n$ die Varietät $\mathcal{V}_k = \overline{V(f_1, \ldots, f_k) \setminus V(g)}$ sich in Noether–Normalform der Dimensionen $r := n - k$ befindet, $0 \in \mathbb{k}^r$ eine Lifting–Faser ist und X_{r+1} ein primitives Element sowohl der Noether–Normalform von $\mathcal{V}_k$ als auch der Faser $\mathcal{V}_k^0$ über $0 \in \mathbb{k}^r$ ist.

Sei $k \in \{1, \ldots, n-1\}$ fixiert und $r := n - k$. Es sei vorausgesetzt, dass die geometrische Lösung der Faser $\mathcal{V}_k^0$ über $0 \in \mathbb{k}^r$ von $\mathcal{V}_k$ schon bekannt ist. Aus dieser soll die Faser $\mathcal{V}_{k+1}^0$ über $0 \in \mathbb{k}^{r-1}$ von $\mathcal{V}_{k+1}$ und damit implizit auch die Parametrisierung à la Kronecker von $\mathcal{V}_{k+1}$ bestimmt werden. In der Darstellung dieses Schritts folgen wir wieder [GLS01], s. auch [HMW01, Leh99, Lec01b].

Damit das hier dargestellte Verfahren durchführbar ist, muss vorausgesetzt werden, dass das Polynom f_{k+1} auf keiner der irreduziblen Komponenten identisch verschwindet. Für das zu vermeidende Polynom g des Systems $\mathcal{S}_k$ ist dies schon nach Konstruktion der Fall. Unter diesen Voraussetzungen gibt es Fasern der Normalform von $\mathcal{V}_k$ auf welchen weder f_{k+1} noch g verschwinden. Diese Bedingung kann wieder als Vermeiden der Nullstellenmenge eines Polynoms in den Koordinaten des Lifting–Punkts ausgedrückt werden. Seien diese Bedingungen zu den anderen Anforderungen an den generischen affin–linearen Koordinatenwechsel hinzugefügt. Damit kann angenommen werden, dass weder f_{k+1} noch g in den Punkten der Faser $\mathcal{V}_k^0$ verschwinden.

Seien $q, v_{r+2}, \ldots, v_n \in \Bbbk[Y]$ die Polynome der geometrischen Lösung

$$q(Y) = 0, \quad \begin{cases} x_m = 0, & m = 1, \ldots, r, \\ x_{r+1} = Y, & \\ q'(Y)x_m = v_m(Y), & m = r+2, \ldots, n, \end{cases}$$

der Faser von $\mathcal{V}_k$ über dem Nullpunkt. Mittels des Newton–Hensel–Verfahrens wird aus dieser Faser die eindimensionale Kurvenschar gewonnen, die die Lösungsmenge des um die Bedingung $x_r = 0$ reduzierten Systems $\mathcal{S}_k''$:

$$x \in \overline{\Bbbk}^n : \quad f_1(x) = \cdots = f_k(x) = x_1 = \cdots = x_{r-1} = 0, \quad g(x) \neq 0$$

ist. Dabei wird aus der geometrischen Lösung der Faser eine Parametrisierung mit Polynomen $Q, V_{r+2}, \ldots, V_n \in \Bbbk[T][Y]$ und $Q' := \frac{\partial Q}{\partial Y}$ der eindimensionalen Varietät $\mathcal{V}_k^L$ des Systems $\mathcal{S}_k''$ erzeugt, diese Parametrisierung hat die Form

$$Q(T)(Y) = 0, \quad \begin{cases} x_m = 0, & m = 1, \ldots, r-1, \\ x_r = T, & \\ x_{r+1} = Y, & \\ Q'(T)(Y)x_m = V_m(T)(Y), & m = r+2, \ldots, n\,. \end{cases}$$

Diese Kurvenschar enthält die Faser $\mathcal{V}_{k+1}^0$ in der Schnittmenge mit der Hyperfläche $V(f_{k+1})$. Sei zur Bestimmung dieser Schnittmenge die rationale Funktion

$$F := f_{k+1}\left(0, \ldots, 0, T, Y, \frac{V_{r+2}(T,Y)}{Q'(T,Y)}, \ldots, \frac{V_n(T,Y)}{Q'(T,Y)}\right) \in \Bbbk(T,Y)$$

definiert, in welcher die abhängigen Variablen in f_{k+1} durch ihre rationalen Ausrücke in $\Bbbk(T,Y)$ ersetzt wurden. Da der Nenner von F eine Potenz von $Q'(T,Y)$ ist und $Q'(T,Y)$ in $\Bbbk(T)[Y]$ modulo $Q(T,Y)$ invertierbar ist, kann F so erweitert und um Vielfache von Q modifiziert werden, dass schon $F \in \Bbbk(T)[Y]$ gilt.

Die gleiche Konstruktion kann für das zu vermeidende Polynom g vorgenommen werden, sei $G \in \Bbbk(T)[Y]$ das entstehende Polynom. Es sind nun diejenigen Paare $(x_r, x_{r+1}) \in \overline{\Bbbk}^2$ zu finden, welche Lösungen des Systems

$$Q(x_r, x_{r+1}) = F(x_r, x_{r+1}) = 0 \quad \text{und} \quad G(x_r, x_{r+1}) \neq 0$$

sind. Sei

$$R(T) := \operatorname*{Res}_Y(Q(T,Y), F(T,Y)) \in \Bbbk(T)$$

die Resultante nach Y. Es läßt sich zeigen, dass R schon ein Polynom in $\Bbbk[T]$ ist, d.h. mit Nenner 1 darstellbar ist. Weiter ergibt sich, dass der Grad dieses Polynoms durch das Produkt der Grade von Q und f_{k+1} beschränkt ist (s. [GLS01], auch [Leh99]).

Da kein Punkt der Faser $\mathcal{V}_k^0$ eine Nullstelle von f_{k+1} ist, muss der konstante Koeffizient der Resultante R von Null verschieden sein. Damit ist gesichert, dass F und Q keine gemeinsamen Faktoren besitzen.

Nach Voraussetzung der generischen affin–linearen Transformation trennt X_r die Lösungen des Systems $Q = F = 0$, $G \neq 0$. Die generische affin–lineare Transformation läßt sich weiter so wählen, dass X_r alle gemeinsamen Nullstellen von Q und F trennt, d.h. auch jene gemeinsamen Nullstellen, die zu Punkten in der Hyperfläche $V(g)$ gehören.

Für jeden Punkt $x = (x_1, \ldots, x_n) \in \mathcal{V}_{k+1}^0$ ist $(x_r, x_{r+1}) \in \overline{\Bbbk}^2$ eine gemeinsame Nullstelle von $Q(T,Y)$ und $F(T,Y)$. Die Resultante $R(T) := \operatorname{Res}_Y(Q,F)$ hat somit in x_r eine Nullstelle. Ist $z \in \Bbbk$ beliebig und $U := T + zY$, so ist auch $(x_r + zx_{r+1}, x_{r+1}) \in \overline{\Bbbk}^2$ eine gemeinsame Nullstelle von $F_z(U,Y) := F(U - zY, Y)$ und $Q_z(U,Y) := Q(U - zY, Y)$. Sind für zwei verschiedene Parameter $z_1, z_2 \in \Bbbk$ die Werte für $x_r + zx_{r+1}$ und $x_r + zx_{r+1}$ bekannt, so können daraus x_r, x_{r+1} und damit alle anderen Koordinaten zurückgewonnen werden. Das zu lösende Problem ist nun, die Nullstellen der Resultanten von einerseits Q_{z_1} und F_{z_1} und andererseits von Q_{z_2} und F_{z_2} einander zuzuordnen.

Da vorausgesetzt wurde, dass X_r die Punkte von $Q = F = 0$ trennt, genügt es, mit einem infinitesimalen Parameter zu arbeiten. Dann sind die Nullstellen der jeweiligen Resultanten dadurch zuzuordnen, dass zusammengehörige Nullstellen sich nur infinitesimal unterscheiden. Sei $\varepsilon := Z + \langle Z^2 \rangle \in \Bbbk[Z]/\langle Z^2 \rangle$, dies ist ein algebraisches Infinitesimal, denn es gilt $\varepsilon^2 = 0$. Dann ist

$$Q_\varepsilon(U,Y) := Q(U - \varepsilon Y, Y) = Q(U,Y) - \varepsilon \frac{\partial Q}{\partial U}(U,Y)Y$$

und $Q'_\varepsilon(U,Y) := \frac{\partial Q_\varepsilon}{\partial Y}(U,Y)$ ist in $\Bbbk(U)$ modulo $Q_\varepsilon(U,Y)$ invertierbar. $F_\varepsilon(U,Y) := F(U - \varepsilon Y, Y)$ kann somit wieder modulo $\langle Q_\varepsilon \rangle$ als polynomial in Y mit rationalen Koeffizienten angenommen werden, $F_\varepsilon \in \Bbbk(U)[Y]$. Sei

$$R_\varepsilon(U) := \operatorname*{Res}_Y(Q_\varepsilon, F_\varepsilon) \in \Bbbk(U)\,.$$

Aus den Eigenschaften der Noether–Normalform folgt wieder, dass R_ε ein Polynom in U ist. Es gibt also Polynome $a_0, a_1 \in \Bbbk[U]$ mit

$$R_\varepsilon(U) = a_0(U) + \varepsilon a_1(U)\,.$$

Nach Rücksubstitution von $U = T + \varepsilon Y$ folgt

$$R_\varepsilon(T + \varepsilon Y) = a_0(T) + \varepsilon\big(a_0'(T)Y + a_1(T)\big).$$

Weiterhin folgt, dass es in der Faktorisierung von $\mathbb{R}_\varepsilon$ keinen irreduziblen Faktor gibt, der unter Auswertung in $\varepsilon = 0$ weiter reduzierbar wäre. Sei $m := ggT(a_0, a_0')$ derjenige Faktor von a_0, mit welchem der verbleibende Faktor $\tilde{a}_0 := a_0/m$ quadratfrei ist. Dann ist m ebenfalls ein Teiler von a_1. Sei $\tilde{a}_1 := a_1/m$. Das Polynom $p := a_0'/m$ ist dann modulo $\tilde{a}_0$ invertierbar, und mit der Lösung $\tilde{w} \in \Bbbk[T]$ von $p\tilde{w} = \tilde{a}_1 \bmod \tilde{a}_0$ ist

$$\tilde{a}_0(T) = 0 \begin{cases} x_r = T \\ x_{r+1} = \tilde{w}(T) \end{cases}$$

eine geometrische Lösung des Systems $Q = F = 0$.

Um zur Lösung von V_{k+1} zu gelangen, muss aus $\tilde{a}_0$ noch der gemeinsame Faktor mit dem Zähler von $G(T, \tilde{w}(T))$ entfernt werden. Sei $q \in \Bbbk[T]$ der verbleibende Faktor, und sei $w_{r+1} \in \Bbbk[T]$ der gradminimale Repräsentant von $\tilde{w} + \langle\, q \,\rangle$. Die so erhaltene geometrische Lösung

$$\tilde{q}(T) = 0 \begin{cases} x_r = T \\ x_{r+1} = w_{r+1}(T) \end{cases}$$

des Systems $Q = F = 0$, $G \neq 0$ kann nun in die Parametrisierung der Kurvenschar $\mathcal{V}_k^L$ eingesetzt werden. Es ist leicht nachzuprüfen, dass $\frac{\partial Q}{\partial Y}(T, w_{r+1}(T)) \in \Bbbk[T]$ modulo $q(T)$ invertierbar ist.

2.5.4 Das Berechnungsmodell „arithmetisches Netzwerk“

Wir können innerhalb eines Polynomrings $\mathcal{R}[\underline{X}] := \mathcal{R}[X_1, \ldots, X_n]$ diejenige Teilmenge betrachten, deren Elemente sich aus den erzeugenden Polynomen $X_1, \ldots, X_n$ und den Ringelementen *konstruieren* lassen, d.h. wir betrachten

- alle konstanten Polynome mit Wert in $\mathcal{R}$ als *konstruierbar*,
- die Variablen $X_k : \mathcal{R}^n \to \mathcal{R}$ als *konstruierbar*,
- die Summe $f + g$ zweier konstruierbarer Polynome $f, g \in \mathcal{R}[\underline{X}]$, als *konstruierbar*, und
- das Produkt $f \cdot g$ zweier konstruierbarer Polynome $f, g \in \mathcal{R}[\underline{X}]$, als *konstruierbar*.

Nach dieser Definition hat jedes konstruierbare Polynom eine Entstehungsgeschichte, d.h. es gibt eine Vorschrift, nach welcher es aus den Konstanten des Rings und den elementaren Polynomen $X_1, \ldots, X_n$ konstruiert werden kann. Diese Vorschrift kann durch einen gerichteten Graphen repräsentiert werden, dessen Knoten Ringoperationen und dessen Kanten Polynome repräsentieren (s. [Mor97] und dort zitierte Literatur, insb. [Kal85]):

Definition 2.5.3 *Ein* gerichteter Graph *ist ein Paar (V, E) einer endlichen Knotenmenge V und einer Kantenmenge $E \subset V \times V$, d.h. jede Kante ist ein geordnetes Paar von Knoten. Der erste Knoten wird als* Anfang, *der zweite als* Ende *der Kante bezeichnet. Ein* Kantenzug *ist eine endliche Folge von Kanten, so dass der Endknoten jeder Kante der Anfangsknoten der nachfolgenden Kante ist. Ein Graph heißt* azyklisch, *wenn es keinen Kantenzug gibt, in welchem der Anfangsknoten der ersten Kante auch der Endknoten der letzten Kante ist. Die* Tiefe *eines azyklischen gerichteten Graphen ist die maximale Anzahl an Kanten in allen möglichen Kantenzügen des Graphen.*

Eine Bewertung *der Knoten eines Graphen ist eine Abbildung mit der Knotenmenge V als Definitionsbereich.*

Zwei elementare Bewertungen mit Werten in den natürlichen Zahlen sind der *Eingangs– und Ausgangsgrad*. Diese ordnen jedem Knoten im Graph jeweils die Anzahl an Kanten zu, für welche dieser Knoten der End– bzw. Anfangsknoten ist.

Definition 2.5.4 *Ein* arithmetisches Netzwerk *ist ein gerichteter azyklischer Graph, dessen Knoten nur Eingangsgrade* 0 *oder* 2 *aufweisen und auf dessen Knotenmenge eine Bewertung mit Werten in der Menge* $\mathcal{R} \cup \{X_1, \ldots, X_n\} \cup \{+, \cdot\}$ *definiert ist.*

Dabei haben die Knoten mit Eingangsgrad 0 *ausschließlich Bewertungen im Ring oder in den Variablen. Ein mit einem Element von* $\mathcal{R}$ *bewerteter Knoten repräsentiert das konstante Polynom mit dieser Konstanten als Wert. Ein mit einer Variablen* X_k *bewerteter Knoten repräsentiert das aus dieser bestehende Polynom und wird* Eingangsknoten *genannt. Es sei zu jeder Variablen genau ein Eingangsknoten vorhanden.*

Knoten mit Eingangsgrad 2 *haben ausschließlich Bewertungen in den Ringoperationen „*$+$*“ und „*$\cdot$*“, wobei ein solcher Knoten die Summe oder das Produkt der eingehenden Kanten, genauer der zwei Polynome an deren Anfang, repräsentiert.*

Zusätzlich zu den Eingangsknoten *wird ein Tupel von* Ausgangsknoten *des Netzwerkes festgelegt.*

Bemerkung: Was hier als arithmetisches Netzwerk definiert wurde, müsste genauer *divisions– und verzweigungsfreies* arithmetisches Netzwerk heißen. Wir werden jedoch nur diesen Typ betrachten, so dass wir bei der kürzeren Bezeichnung bleiben.

Wir können nun die Operationen, welche zur Konstruktion von Polynomen verwendet werden, mittels arithmetischer Netzwerke nachvollziehen. Die erzeugenden Polynome, d.h. die konstanten Polynome und die Variablen, werden durch die einfachsten arithmetischen Netzwerke verkörpert, deren einziger Ausgangsknoten Eingangsgrad 0 hat. Nach Definition ist dieser Ausgangsknoten entweder mit einer Variablen oder einer Ringkonstanten bewertet.

Mehrere Netzwerke können zu einem einzigen zusammengefasst werden, indem ihre Ecken– und Kantenmengen vereinigt und identische Eingangsknoten zusammengefasst werden. Ebenso kann ein Netzwerk mit mehreren Ausgangsknoten aufgespalten werden, indem von ihm Kopien angefertigt werden und die Menge der Ausgangsknoten nach Bedarf verkleinert wird. Bei beiden Operationen können die Netzwerke im folgenden Sinn optimiert werden. Bei der Vereinigung können Teilgraphen mit gemeinsamen Eingangsknoten und identischen Verknüpfungen dieser identifiziert werden, beim Aufspalten können Knoten entfernt werden, welche nicht mittels eines Kantenzuges mit einem der Ausgangsknoten verbunden sind.

Zwei Netzwerke, welche jedes ein Polynom repräsentieren, werden durch Addition oder Multiplikation verknüpft, indem sie zu einem Netzwerk zusammengefasst werden und ihre Ausgangsknoten mittels neu hinzugefügter Kanten mit einem ebenso neuen Ausgangsknoten verknüpft werden. Dieser neue Ausgangsknoten wird entsprechend der Operation mit „$+$“ oder „$\cdot$“ bewertet.

Ist ein Polynom als Ausgangsknoten eines arithmetischen Netzwerks gegeben, so kann, ausgehend von den Eingangsknoten, zu jedem Knoten das repräsentierte Polynom in Monomdarstellung bestimmt werden, so dass man nach endlich vielen Schritten das gegebene Polynom auch in Monomdarstellung vorliegen hat. Umgekehrt kann jedem in Monomdarstellung gegebenen Polynom ein arithmetisches Netzwerk zugeordnet werden, indem die Monome durch ein Netzwerk einfacher Multiplikationen dargestellt werden und dann die Summe der Monome gebildet wird. Somit sind alle Polynome in $\mathcal{R}[X_1, \ldots, X_n]$ durch ein arithmetisches Netzwerk über $\mathbb{R}$ konstruierbar.

Die Zuordnung einer Monomdarstellung zu einem arithmetischen Netzwerk ist eindeutig, jedoch können demselben Polynom viele verschiedene Konstruktionsvorschriften zugeordnet werden. Dies mag als Nachteil der Netzwerkdarstellung erscheinen. Ein Vorteil dieser ergibt sich aber daraus, dass sich aus vielen praktischen oder geometrischen Aufgaben ein kurzes arithmetisches Netzwerk für die auftretenden Polynome schon aus der Aufgabenstellung ableiten läßt. Dieses Netzwerk wird meist kürzer sein, als ein aus der Monomdarstellung abgeleitetes.

Ein arithmetisches Netzwerk kann mit nur wenigen Knoten und Kanten Polynome hohen Grades kodieren, deren Koeffizientenfolge aber dicht besetzt ist. Ein künstliches Beispiel ist das Polynom $\prod_{k=1}^{N}(k + (X_1 + X_2 + \cdots + X_n))$, welches sich durch ein Netzwerk mit $n + N$ Additions– und N Multiplikationsknoten darstellen läßt. Es hat jedoch einen Grad N und ist dicht besetzt, die Koeffizientenfolge enthält also $\binom{N+n}{n} \approx N^n$ Monome. Beachtet man, dass meist die Multiplikation im Ring wesentlich aufwendiger als die Addition auszuführen ist, so erkennt man in diesem Beispiel einen wesentlichen Unterschied beider Darstellungsformen, sowohl in Betracht der Kodierung, als auch der Auswertung.

2.5.5 Komplexitätsmaße arithmetischer Netzwerke

Arithmetische Netzwerke lassen sich als *Algorithmen* betrachten, welche als Unterprogramme auf einem Computer ausgeführt werden können. Wir müssen dazu annehmen, dass dieser Computer die Ringelemente darstellen und die Ringoperationen ausführen kann. Man kann nun zwischen *seriellen* und *parallelen* Computern unterscheiden, je nach Anzahl der gleichzeitig ausführbaren Ringoperationen. Dabei vertritt der serielle Computer den Extremfall, dass nur eine Ringoperation nach der anderen ausgeführt werden kann, der parallele Computer hingegen den anderen Extremfall, dass beliebig viele Rechenoperationen gleichzeitig ausgeführt werden können, sofern alle Operanden bekannt sind.

Die parallele Ausführung eines arithmetischen Netzwerkes kann nach der Höhe der Knoten organisiert werden. Dabei ist die Höhe eine rekursiv definierte Bewertung der Knotenmenge, wobei den Knoten des Eingangsgrades 0, also Variablen und Ringelementen, die Höhe 0 zugeordnet wird, jedem weiteren Knoten wird das um Eins erhöhte Maximum der Höhen der Vorgängerknoten zugeordnet. Anders ausgedrückt, die Höhe eines Knotens ist die maximale Anzahl von Kanten in einem Kantenzug, der diesen als letzten Knoten besitzt. Alle Knoten mit gleicher Höhe werden gleichzeitig im Zeitschritt, der ihrer Höhe entspricht, ausgewertet.

Um eine eindeutig definierte Auswertung des Netzwerkes auf einem seriellen Computer zu fixieren, muss eine Berechnungsreihenfolge, ein sog. „*Game of Pepples*“, festgelegt werden.

Definition 2.5.5 *Seien Polynome $f_1, \ldots, f_p \in \mathcal{R}[X_1, \ldots, X_n]$ durch die Ausgangsknoten eines arithmetischen Netzwerkes $\mathcal{N}$ gegeben. Ein „*Pebblegame*“ auf $\mathcal{N}$ ist eine Ordnung auf der Knotenmenge dieses Netzwerkes, in welcher für jede Kante der Anfangsknoten kleiner als der Endknoten ist. Ein arithmetisches Netzwerk mit Pebblegame wird auch als* Straight–Line–Program *(SLP) bezeichnet.*

Durch ein „Pebblegame“, d.h. eine Ordnung der Knotenmenge, wird eine Auswertungsvorschrift des arithmetischen Netzwerkes definiert: Die Werte der Polynome $f_1, \ldots, f_p$ in einem Punkt $x = (x_1, \ldots, x_n) \in \mathcal{R}^n$ erhält man, indem man die Knoten in der durch die Ordnung des Pebblegame gegebenen Reihenfolge durchläuft und dabei auswertet. Die als konstant bewerteten Knoten erhalten ihr Ringelement als Wert, die Koordinatenknoten das entsprechende Glied des Argumenttupels x. Den Operationsknoten wird das Ergebnis der Auswertung der Operation des Knotens, wobei die Operanden die Werte der zwei Vorgängerknoten sind. Nach Definition des Pebblegame sind diese Vorgängerknoten bereits ausgewertet. Das Netzwerk ist ausgewertet, wenn alle Ausgangsknoten ausgewertet wurden. Der Wert des Netzwerkes ist dann das Tupel der Ringelemente, die sich als Werte in den Ausgangsknoten ergeben.

Definition 2.5.6 *Die* Komplexität *eines arithmetischen Netzwerkes im parallelen Modus (d.h. ohne Pebblegame) wird als Paar (ℓ, L) aus* paralleler Höhe und Breite *angegeben. Jedem Knoten wie oben beschrieben eine Höhe zugeordnet. Die* parallele Höhe ℓ *ist die maximale Höhe eines Ausgangsknotens und die* parallele Breite L *ist die maximale Anzahl von Knoten einer gemeinsamen Höhe.*

Oft wird die Bestimmung der Höhe eines Knotens auch so eingerichtet, dass diese nur von nichttrivialen Multiplikationen erhöht wird. D.h. einer Addition oder einer Multiplikation mit einer Ringkonstanten wird die maximale Höhe der Operanden zugeordnet, nur im verbleibenden Fall, bei der sog. *nichtskalaren* Multiplikation, wird die maximale Höhe der Operanden um Eins erhöht. Dementsprechend heißen die daraus abgeleiteten Kenngrößen des Netzwerkes *nichtskalare parallele Höhe* und *nichtskalare parallele Breite*. Die Höhe, insbesondere die nichtskalare Höhe ℓ liefert eine Schranke 2^ℓ für die Grade der durch die Ausgangsknoten verkörperten Polynome. Auf einem idealen Computer, welcher beliebig viele, aber mindestens L Berechnungseinheiten für Operationen im Grundring hat, würde die parallel organisierte Auswertung nur ℓ Berechnungsschritte benötigen.

Definition 2.5.7 *Die Komplexität eines* Straight–Line–Programms, *d.h. eines arithmetischen Netzwerks mit Pebblegame, wird als Paar (T, S) aus* seriellem Zeit– und Platzbedarf *angegeben. Dabei ist der* serielle Zeitbedarf T *die Anzahl der Knoten, d.h. der Rechenoperationen. Der* serielle Platzbedarf S *ist definiert als das Maximum des Platzbedarfs der einzelnen Knoten. Der Platzbedarf eines Knotens ist die Anzahl der vor diesem Knoten schon bestimmten, aber nach diesem Knoten noch benötigten Zwischenergebnisse. Genauer können zu jedem Knoten Vorgänger und Nachfolger bzgl. der Ordnung des Pebblegames definiert werden. Dabei zähle jeder Knoten selbst zu seinen Vorgängern. Die Anzahl der Vorgänger eines Knotens, die durch Kantenzüge mit Nachfolgern verbunden sind, ist der Platzbedarf dieses Knotens.*

Verschiedene Pebblegames zum selben Netzwerk können verschiedene Komplexitäten der seriellen Auswertung ergeben.

2.5.6 Zum Vergleichen arithmetischer Netzwerke

Wie schon angesprochen, können verschiedene arithmetische Netzwerke dasselbe Polynom verkörpern. Um Netzwerke zu vergleichen, d.h. festzustellen, ob sie als Polynome identisch sind, ist es also nicht ausreichend, die Struktur ihrer Graphen zu vergleichen.

Man könnte die Monomdarstellungen der Netzwerke bestimmen und diese vergleichen. Im Verlaufe dieses Verfahrens müssten im Allgemeinen etwa $2^{\ell n}$ Koeffizienten bestimmt werden, dabei ist ℓ die Höhe des arithmetischen Netzwerks und n die Anzahl der Unbestimmten im Polynomring. Ein solcher exponentiell in den Eingangsparametern wachsender Aufwand wird als unpraktikabel angesehen.

Es ist möglich, die Gleichheit zweier arithmetischer Netzwerke bzw. die Identität eines Netzwerks zum Nullpolynom ohne Bestimmung der Monomkoeffizienten durch mehrfaches Auswerten des Netzwerkes an hinreichend vielen Stellen zu prüfen.

Definition 2.5.8 (vgl. [Zip79, HS80a, GHMP95]) *Sei $\mathbb{K}$ ein Körper und $\mathcal{F}$ eine Teilmenge von Polynomen aus $\mathbb{K}[X_1, \ldots, X_n]$. Eine Teilmenge $\mathcal{Q} \subset \mathbb{K}^n$ heißt* korrekte Testfolge *oder* Questor*, wenn für beliebige Polynome $f \in \mathcal{F}$ gilt*

$$\textit{Wenn } f(x) = 0 \textit{ für alle } x \in \mathcal{Q}, \textit{ dann ist schon } f = 0.$$

Die Mächtigkeit von $\mathcal{Q}$ heißt Länge *des Questors.*

Die Existenz und Größe von Questoren für arithmetische Netzwerke läßt sich allein über ihre nichtskalare Höhe und Breite charakterisieren.

Satz 2.5.9 (Zippel–Schwartz Test, [Zip79, HS80a, Sch80] nach [GHMP95])
Sei $\mathcal{F}$ die Familie von Polynomen in $\mathbb{K}[X_1, \ldots, X_n]$, welche sich durch ein arithmetisches Netzwerk der nichtskalaren Breite L und Höhe ℓ verkörpern lassen. Seien $\omega := 2(2^\ell - 1)(2^\ell + 1)^2$ und $\sigma := 6(\ell L)^2$. Sei weiter $\Omega \subset \mathbb{K}$ eine Teilmenge mit ω Elementen. Unter allen $\binom{\omega^n}{\sigma}$ endlichen Folgen mit σ Elementen in $\Omega^n \subset \mathbb{K}^n$ gibt es mindestens $\omega^{n\sigma}(1 - \omega^{-\frac{\sigma}{6}})$ korrekte Testfolgen.

Wenn man also die Menge $\Omega := \{-\frac{\omega}{2}, \ldots, \frac{\omega}{2}\} \subset \mathbb{Z}$ vorgibt und aus der Menge Ω^n der n–Tupel zufällig eine „gute“ Folge von σ Elementen auswählt, so besagt das Theorem, dass jedes Polynom, welches in allen Tupeln der Folge eine Nullstelle hat, schon das Nullpolynom sein muss. Die Fehlerwahrscheinlichkeit, dass die gewählte Folge nicht diese „gute“ Eigenschaft hat, ist kleiner als $\omega^{-\sigma/6}$. Diese Schranke ist für realistische Größen von ℓ und L eine sehr kleine Zahl.

Die für die Komplexität dieses Verfahrens entscheidende Anzahl von maximal $\sigma := 6(\ell L)^2$ Auswertungen des Netzwerks ist polynomial in den Parametern des Netzwerks und ist damit als praktikabel anzusehen. Die zweite Größe $\omega \leq 2^{3(\ell+1)}$ ergibt eine linear in ℓ wachsende Bitlänge der Testpunkte. Meist ergibt sich schon nach sehr viel weniger als σ Auswertungen, dass das betrachtete Polynom vom Nullpolynom verschieden ist.

Die zufällige Natur von auf arithmetischen Netzwerken beruhender Algorithmen besteht nun gerade darin, dass in einem solchen Test ein von Null verschiedenes Polynom falsch als Null identifiziert werden kann. Dies wird jedoch im Zusammenhang des umgebenden Algorithmus meist zu einem Abbruch, z.B. wegen versuchter Division durch Null, oder im weiteren Verlauf auf eine gut erkennbare Abweichung führen. Damit werden solche Algorithmen entweder mit einem gültigen Ergebnis oder einem Abbruch unbestimmter Natur enden. Durch mehrere Läufe mit verschiedenen Testfolgen kann die Wahrscheinlichkeit eines Abbruchs aufgrund eines schlecht gewählten Questors beliebig weit reduziert werden.

2.5.7 Komplexitätsabschätzungen des TERA–Kronecker–Verfahrens

In diesem Abschnitt sollen Komplexitätsabschätzungen für die Lösung polynomialer Systeme zusammengetragen werden, die in den Arbeiten [HMW01, GLS01, Lec01b] erhalten wurden. Zunächst seien die Anforderungen an ein „gutes“ Gleichungssystem formuliert, für welches die bisher in diesem Abschnitt dargestellten Algorithmen anwendbar sind. Die dazu geforderten Einschränkungen des allgemeinen Falls können mit entsprechenden Modifikationen des Algorithmus überwunden werden, wie in [Lec01b] ausgeführt wurde, (die dazu wesentlichen Abschnitte wurden in [Lec01a, Lec03] publiziert).

Definition 2.5.10 *Seien $\Bbbk$ ein Körper der Charakteristik 0, $f_1, \ldots, f_n \in \Bbbk[X_1, \ldots, X_n]$ Polynome.*

Die Folge $(f_1, \ldots, f_n)$ heißt

- regulär, *wenn für jedes $k = 1, \ldots, n-1$ das Polynom f_{k+1} auf keiner irreduziblen Komponente von $V(f_1, \ldots, f_k)$ verschwindet. D.h. die Restklasse von f_{k+1} im Restklassenring $\Bbbk[X_1, \ldots, X_n]/\langle f_1, \ldots, f_k \rangle$ ist kein Nullteiler.*
- reduziert, *wenn für jedes $k = 1, \ldots, n-1$ das Ideal $\langle f_1, \ldots, f_k \rangle$ radikal ist und*
- transversal, *wenn sie sowohl regulär als auch reduziert ist.*

Diese Definition kann auch auf ein allgemeines System mit einer zu vermeidenden Hyperfläche $V(g)$ erweitert werden.

Sei eine transversale Folge $f_1, \ldots, f_n \in \Bbbk[X_1, \ldots, X_n]$ von Polynomen in $n \in \mathbb{N}$ Variablen gegeben. Fügen wir als weitere Unbestimmte die Einträge $A_{1,1}, \ldots, A_{n,n}$ einer $n \times n$ Matrix und $P_1, \ldots, P_n$ eines Punktes P der Dimension n hinzu. Sei $\mathbb{K} := \Bbbk(A, P)$ der Körper der rationalen Funktionen über diesen Unbestimmten und $\overline{\mathbb{K}}$ ein algebraischer Abschluss von $\mathbb{K}$. Dann ist auch die Folge $F_1, \ldots, F_n \in \mathbb{K}[Y_1, \ldots, Y_n]$ mit $F_k(A, P)(Y) := f_k(P + AY)$ transversal. Nach [GH91] (Abschn. „Methode brutale“) befinden sich dann alle Varietäten $\mathcal{V}_k := V(F_1, \ldots, F_k) \subset \overline{\mathbb{K}}^n$, $k = 1, \ldots, n$, in Noether–Normalform der Dimension $r := n - k$, die Faser $\mathcal{V}_k^0$ über $0 \in \Bbbk^r$ ist eine Lifting–Faser, X_{r+1} ist primitives Element etc.

Jedoch ist das Rechnen im Körper der rationalen Funktionen bzgl. der Komplexität recht ungünstig. Es ist also vorzuziehen, die Parameter in einem generischen Punkt auszuwerten, d.h. eine Matrix $A \in \Bbbk^{n \times n}$ und einen Punkt $P \in \Bbbk^n$ zu wählen, mit welchen die geforderten Eigenschaften immer noch erfüllt sind.

Satz 2.5.11 (simultane Noether–Normalisierung, s. [HMW01], Thm. 3)
Sei $f_1, \ldots, f_n \in \mathbb{Q}[X_1, \ldots, X_n]$ eine transversale Folge. Dann gibt es eine ganzzahlige Matrix $A \in \mathbb{Z}^{n\times n}$ und einen Punkt $P \in \mathbb{Z}^n$, so dass für jedes $k = 1, \ldots, n$ und $r := n - k$ gilt:

i) Die Varietät $\mathcal{V}_k := V(f_1^A, \ldots, f_k^A) \subset \mathbb{C}^n$ befindet sich in Noether–Normalform der Dimension r.

ii) Der Punkt $P^{(r)} := \pi_r(P) = (p_1, \ldots, p_r) \in \mathbb{Z}^r$ ist ein Lifting–Punkt für $\mathcal{V}_k$.

iii) Ist $k < n$, so sind die Fasern $\mathcal{V}_k \cap \pi_r^{-1}(\pi_r(\xi))$ regulär für $\xi \in V_{k+1} \cap \pi_{r-1}^{-1}(P^{(r-1)})$.

Seien $\delta = \max\{\deg V(f_1, \ldots, f_k) : k = 1, \ldots, n\}$ der maximale geometrische Grad der Zwischenvarietäten, $d \in \mathbb{N}$ eine Schranke der Grade der Polynome $f_1, \ldots, f_n$ und sei $\kappa \in \mathbb{N}_{>0}$ beliebig. Dann können die Einträge von A und P im Bereich $\{1, \ldots, 8\kappa n^8 d^4 \delta^9\}$ zufällig gewählt werden. Die Wahrscheinlichkeit, ein geeignetes Paar (A, P) zu erzeugen, ist größer als $(1 - \frac{1}{\kappa})^2 \geq \frac{1}{4}$.

Der in [HMW01] untersuchte Algorithmus verwendet als primitive Elemente zufällig gewählte Linearkombinationen der Variablen mit ganzzahligen Koeffizienten mit einer Schranke $\kappa\delta^2$ des Absolutbetrages. Da sich die Eignung als primitives Element von einer gegebenen Linearkombination direkt im Laufe des Algorithmus überprüfen läßt, ist es nicht notwendig, diese in die simultane Noether–Normalisierung mit einzubeziehen. Zur Komplexität des gesamten Verfahrens ergibt sich die folgende Aussage.

Satz 2.5.12 ([HMW01], Thm. 1) *Seien die Voraussetzungen wie im vorherigen Satz, zusätzlich gebe es zu den Polynome $f_1, \ldots, f_n \in \mathbb{Q}[X_1, \ldots, X_n]$ ein Straight–Line–Programm mit Zeit– und Platzbedarf (T, S), welches diese Polynome auswertet. Dann kann eine geometrische Lösung des Systems*

$$x \in \mathbb{C}^n : \quad f_1(x) = \cdots = f_n(x) = 0$$

mittels eines probabilistischen Algorithmus mit

$$\begin{aligned} \textit{Platzbedarf} \quad & O(Sdn\delta^2) \\ \textit{und Laufzeit} \quad & O\Big((Tdn^2 + n^5)\delta^3(\log_2 \delta)^2(\log_2 \log_2 \delta)^2\Big) \end{aligned}$$

bestimmt werden.

Dieses Resultat wurde im Verlaufe der Implementierung des TERA–Kronecker–Verfahrens ([Lec01b]) weiter verbessert.

Satz 2.5.13 ([GLS01], Thm. 1) *Seien $\Bbbk$ ein Körper der Charakteristik 0 und $f_1, \ldots, f_n, g \in \Bbbk[X_1, \ldots, X_n]$ Polynome mit maximalem Grad d, die durch ein Straight–Line–Programm mit Zeitbedarf T gegeben sind. Die Folge $(f_1, \ldots, f_n)$ sei transversal außerhalb $V(g)$.*

Sei δ das Maximum der geometrischen Grade der Zwischenvarietäten $\mathcal{V}_k := \overline{V(f_1, \ldots, f_k) \setminus V(g)}$, $\Omega < 4$ bezeichne den Exponenten der Komplexität von Matrixoperationen über $\Bbbk$, und $M(m) = O(m \log_2(m) \log_2(\log_2(m)))$ bezeichne die Komplexitätsklasse für Operationen mit Polynomen aus $\Bbbk[Y]$ mit Grad höchstens m.

Dann kann eine geometrische Lösung von $V(f_1, \ldots, f_n) \setminus V(g)$ mittels eines probabilistischen Algorithmus mit Laufzeit $O\Big(n(nT + n^\Omega)M(d\delta)^2\Big)$ bestimmt werden.

Die Beendigung des Algorithmus mit einem korrekten Ergebnis hängt von der Wahl einer Anzahl von Parametern in $\mathbb{k}$ ab, die, wie oben in der Beschreibung der einzelnen Schritte des Algorithmus ausgeführt, eine affin–lineare Koordinatentransformation definieren. Die Menge ungeeigneter Parametertupel ist in einer echten algebraischen Varietät des Parameterraums enthalten. Im Falle $\mathbb{k} = \mathbb{Q}$ ist die Menge guter Parameter dicht im Sinne jeder Normtopologie (allgemeiner: Sie ist dicht im Sinne einer $\mathbb{k}$–Zariski–Topologie).

Die Voraussetzung, dass die Polynome $f_1, \ldots, f_n$ eine transversale Folge bilden, kann durch geeignete Erweiterungen des bisher dargestellten Algorithmus fallengelassen werden (s. [Lec03]). Dabei dürfen in den Zwischenvarietäten $\mathcal{V}_k$ Komponenten verschiedener Dimension auftreten. Weiter dürfen irreduzible Komponenten eine algebraische Vielfachheit (vgl. [Mat86]) größer 1 bzgl. des Systems $f_1, \ldots, f_n$ besitzen.

Satz 2.5.14 ([Lec03], Thm. 1) *Seien $\mathbb{k}$ ein Körper der Charakteristik 0 und $f_1, \ldots, f_n, g \in \mathbb{k}[X_1, \ldots, X_n]$ Polynome mit maximalem Grad d, die durch ein Straight–Line–Programm mit Zeitbedarf T ausgewertet werden können. Seien weiter δ der maximale geometrische Grad der Zwischenvarietäten $\mathcal{V}_k := \overline{V(f_1, \ldots, f_k) \setminus V(g)}$ und m die maximale algebraische Vielfachheit einer irreduziblen Komponente der Varietäten $\mathcal{V}_1, \ldots, \mathcal{V}_n$.*

Dann gibt es einen probabilistischen Algorithmus der Laufzeitkomplexität

$$O\big(s \log_2(d) n^4 (nL + n^\Omega) M(d\, m\delta)\big)\ ,$$

der zu jeder äquidimensionalen Komponente der algebraischen Varietät $\mathcal{V}_s$ des Systems

$$f_1 = \cdots = f_s = 0, \quad g \neq 0$$

eine Lifting–Faser bestimmt.

2.6 Analytische Charakterisierung lokaler Extrema

Die Grundlagen für unsere Überlegungen zur polynomialen Optimierung kommen aus der klassischen Analysis. Der Übergang von der Algebra zur Analysis ist leicht gemacht, da Polynome als beliebig oft stetig differenzierbare Funktionen betrachtet werden können. Für die algebraische Interpretation der Ergebnisse der Analysis ist es von fundamentaler Bedeutung, dass die partiellen Ableitungen von multivariaten Polynomen wieder Polynome sind.

2.6.1 Lagrange–Theorie der Extrema

Die Lagrange–Theorie liefert i.a. Kriterien für lokale Extrema einer stetig differenzierbaren Funktion unter Gleichungsnebenbedingungen, welche durch ebenso stetig differenzierbare Funktionen gegeben sind. Die hier zu betrachtenden polynomialen Optimierungsaufgaben sind von dieser Struktur.

Es seien stetig differenzierbare Funktionen $g_1, \ldots, g_p : \mathbb{R}^n \to \mathbb{R}$, $p < n$, gegeben und sei $M := \{x \in \mathbb{R}^n : \ g_1(x) = \cdots = g_p(x) = 0\}$ die nichtleere Menge ihrer gemeinsamen reellen

Nullstellen. Es sei daran erinnert, dass nach Definition 2.5.1 ein Punkt $x \in M$ *regulär* genannt wird, wenn die Jacobi–Matrix

$$J_g(x) := \frac{\partial(g_1,\ldots,g_p)}{\partial(x_1,\ldots,x_n)}(x) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x) & \ldots & \frac{\partial g_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial g_p}{\partial x_1}(x) & \ldots & \frac{\partial g_p}{\partial x_n}(x) \end{pmatrix}$$

den maximalen Rang p hat.

In einem regulären Punkt $x \in M$ enthält die Jacobi–Matrix $J_g(x)$ p linear unabhängige Spalten. Durch Umsortieren der Koordinaten und damit der Spalten läßt sich erreichen, dass die letzten p Spalten linear unabhängig sind.

Nach dem Satz über implizite Funktionen gibt es dann offene Mengen $U \subset \mathbb{R}^{n-p}$ und $V \subset \mathbb{R}^p$, so dass $U \times V$ eine Umgebung von x ist, sowie Funktionen $\varphi_1,\ldots,\varphi_p : U \to \mathbb{R}$, so dass die Umgebung $M \cap (U \times V)$ von x der Graph der Abbildung $\varphi : U \to \mathbb{R}^p$, $\varphi = (\varphi_1,\ldots,\varphi_p)$ ist. D.h. die Abbildung $\Phi : U \to \mathbb{R}^n$ mit

$$z \mapsto \Phi(z) := \big(z, \varphi(z)\big) = \big(z_1,\ldots,z_{n-p}, \varphi_1(z),\ldots,\varphi_p(z)\big)$$

ist injektiv und ihre Bildmenge ist $M \cap (U \times V)$.

Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine weitere stetig differenzierbare Funktion, von welcher wir Minima auf der nichtleeren Menge M suchen. Nach dem Satz von Weierstraß gibt es in den Fällen:

i) M ist kompakt bzw.

ii) f ist nach unten (nach oben) beschränkt,

ein Minimum (Maximum) existiert von f auf M, genauer besitzt jede Zusammenhangskomponente von M ein lokales Minimum (Maximum).

In der *klassischen Lagrange–Theorie* untersucht man nun die Situation, in welcher die lokalen Minima reguläre Punkte von M sind. Sei $x^* \in M$ ein lokaler Minimalpunkt und regulär. Weiter seien die Koordinaten so umgeordnet, dass das System $g_1(z,y) = \cdots = g_p(z,y) = 0$, $z = (x_1,\ldots,x_{n-p})$, $y = (x_{n-p+1},\ldots,x_n)$, nach $y = \varphi(z)$ lokal um x auflösbar ist.

Dann kann $h(z) := f(z,\varphi(z))$ frei von Nebenbedingungen minimiert werden. Da $x^* = (z^*,y^*)$ ein lokales Minimum von f auf M ist, müssen alle partiellen Ableitungen von h in z^* verschwinden. Nach Kettenregel ergibt sich in x^* mit den Matrizen

$$g_z := \frac{\partial(g_1,\ldots,g_p)}{\partial(x_1,\ldots,x_{n-p})}(x^*), \quad g_y := \frac{\partial(g_1,\ldots,g_p)}{\partial(x_{n-p+1},\ldots,x_n)}(x^*),$$
$$f_z := \frac{\partial f}{\partial(x_1,\ldots,x_{n-p})}(x^*), \quad f_y := \frac{\partial f}{\partial(x_{n-p+1},\ldots,x_n)}(x^*)$$
$$\text{und } \varphi_z := \frac{\partial(\varphi_1,\ldots,\varphi_p}{\partial(z_1,\ldots,z_{n-p})}(z^*)$$

das lineare Gleichungssystem

$$0 = g_z + g_y\varphi_z \quad \& \quad 0 = f_z + f_y\varphi_z = f_z - f_y g_y^{-1} g_z\,,$$

welches sich mit den Lagrange–Multiplikatoren $\lambda := -f_y g_y^{-1} \in \mathbb{R}^{1\times p}$ auf eine einheitliche und von Matrixinversionen freie Form bringen läßt,

$$0 = f_y + \lambda g_y \quad \& \quad 0 = f_z + \lambda g_z.$$

Somit kann auf die lokale Auflösung von M wieder verzichtet werden. Es muss lediglich nach solchen Punkten $x \in M$ gesucht werden, in welchen das Differential $Df(x) = J_f(x)$ eine Linearkombination der Differentiale $Dg_1(x), \ldots, Dg_p(x)$ ist. D.h. wir suchen Punkte $x \in M$, in welchen die erweiterte Jacobi–Matrix

$$J_{(g,f)}(x) := \frac{\partial(g_1,\ldots,g_p,f)}{\partial(x_1,\ldots,x_n)}(x) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x) & \ldots & \frac{\partial g_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial g_p}{\partial x_1}(x) & \ldots & \frac{\partial g_p}{\partial x_n}(x) \\ \frac{\partial f}{\partial x_1}(x) & \ldots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

in den ersten p Zeilen linear unabhängig ist und in denen durch die letzte Zeile der Rang nicht erhöht wird. D.h. statt des maximal möglichen Rangs $p+1$ hat die erweiterte Jacobi–Matrix exakt den Rang p.

Definition 2.6.1 *Jeder Punkt $x \in M$, der die Rangbedingung* $\operatorname{rank} J_{(g,f)}(x) = \operatorname{rank} J_g(x) = p$ *erfüllt, wird* kritischer Punkt *für f genannt.*

2.6.2 Minoren der Jacobi–Matrix

Im Falle der polynomialen Optimierung können wir nun weitere Variable $U_1, \ldots, U_p$ einführen und das System

$$\begin{aligned} g_1(X) = \cdots = g_p(X) &= 0 \\ \mathcal{D}_1 f(X) + U_1 \mathcal{D}_1 g_1(X) + \cdots + U_p \mathcal{D}_1 g_p(X) &= 0 \\ &\vdots \\ \mathcal{D}_n f(X) + U_1 \mathcal{D}_n g_1(X) + \cdots + U_p \mathcal{D}_n g_p(X) &= 0 \end{aligned}$$

betrachten. Dies ist ein System von $p+n$ Gleichungen in $n+p$ Variablen. Jedoch ist dieses System in den U_k linear, diese können also mit klassischen Methoden eliminiert werden. Dabei ist es erstrebenswert, die Anzahl der Variablen wie auch der Gleichungen in einem polynomialen Gleichungssystem so klein wie möglich zu halten.

Die Bedingung für kritische Punkte läßt sich ohne Hinzunahme weiterer Variabler durch Minoren der Jacobi–Matrix, d.h. Determinanten von quadratischen Teilmatrizen, ausdrücken: Ein Punkt $x \in \mathbb{R}^n$ ist kritisch für f auf M, wenn

- $g_1(x) = \cdots = g_p(x) = 0$,
- wenigstens eine der folgenden Determinanten von Null verschieden ist

$$\det \frac{\partial(g_1,\ldots,g_p)}{\partial(x_{i_1},\ldots,x_{i_p})}, \quad 1 \le i_1 < i_2 < \cdots < i_p \le n$$

- alle der folgenden Determinanten verschwinden

$$\det \frac{\partial(g_1, \ldots, g_p, f)}{\partial(x_{i_1}, \ldots, x_{i_{p+1}})}, \quad 1 \leq i_1 < i_2 < \cdots < i_{p+1} \leq n \, .$$

Die darin auftretenden partiellen Ableitungen sind, wenn $g_1, \ldots, g_p, f$ Polynome sind, wieder Polynome. Die Determinanten sind polynomial in ihren Einträgen, also auch wieder Polynome. Jedoch entstehen durch diese Bedingungen sehr viele Polynome, welche jedoch nicht voneinander unabhängig sind.

Wir können die Anzahl der zu betrachtenden Matrizen bedeutend reduzieren, indem wir lineare Abhängigkeiten zwischen ihnen in Betracht ziehen.

Lemma 2.6.2 (Austauschlemma, vgl. auch [GH80]) *Seien $\mathbf{a}_1, \ldots, \mathbf{a}_n$, $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1} \in \mathcal{R}^n$ Spaltenvektoren mit Einträgen in einem Ring $\mathcal{R}$, sowie $b_1, \ldots, b_n$, $w_1, \ldots, w_{n+1} \in \mathcal{R}$. Dann gilt*

$$\det(\mathfrak{A}) \det \begin{pmatrix} \mathfrak{V} \\ \mathbf{w} \end{pmatrix} = \sum_{k=1}^{n+1} (-1)^{n+1+k} \det \begin{pmatrix} \mathfrak{A} & \mathbf{v}_k \\ \mathbf{b} & w_k \end{pmatrix} \det \mathfrak{V}_k \, ,$$

wobei $\mathfrak{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ eine $n \times n$–Matrix, $\mathfrak{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_{n+1})$ eine $n \times (n+1)$–Matrix und $\mathfrak{V}_k = (\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_{n+1})$ die um die k–te Spalte reduzierte $n \times n$–Matrix sind, sowie $\mathbf{b} = (b_1, \ldots, b_n)$, $\mathbf{w} = (w_1, \ldots, w_{n+1})$ Zeilenvektoren der Längen n bzw. $n+1$.

Beweis: Nach der Samuelson–Formel, s. Lemma 2.2.18, und nachfolgend der Laplace-Entwicklungsformel für die letzte Zeile einer $(n+1) \times (n+1)$–Matrix, gilt für die rechte Seite

$$\begin{aligned} &\sum_{k=1}^{n+1} (-1)^{n+1+k} \det \begin{pmatrix} \mathfrak{A} & \mathbf{v}_k \\ \mathbf{b} & w_k \end{pmatrix} \det \mathfrak{V}_k = \sum_{k=1}^{n+1} (-1)^{n+1+k} \left(w_k \det \mathfrak{A} - \mathbf{b}\mathfrak{A}^{\#}\mathbf{v}_k \right) \det \mathfrak{V}_k \\ &= \det \mathfrak{A} \det \begin{pmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_{n+1} \\ w_1 & \ldots & w_{n+1} \end{pmatrix} - \det \begin{pmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_{n+1} \\ \mathbf{b}\mathfrak{A}^{\#}\mathbf{v}_1 & \ldots & \mathbf{b}\mathfrak{A}^{\#}\mathbf{v}_{n+1} \end{pmatrix} . \end{aligned}$$

In der letzten Matrix ist die letzte Zeile eine Linearkombination der vorhergehenden Zeilen, somit verschwindet ihre Determinante, es verbleibt die behauptete Identität. □

Lemma 2.6.3 (adaptiert nach [BGHM01]) *Es seien $f, g_1, \ldots, g_p : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, $M := \{g_1 = \cdots = g_p = 0\}$. Gibt es zu einem Punkt $x \in M$ ein Indextupel $I := (i_1, \ldots, i_p)$ mit $1 \leq i_1 < i_2 < \cdots < i_p \leq n$, so dass*

i) $\det \left(\dfrac{\partial(g_1, \ldots, g_p)}{\partial(x_{i_1}, \ldots, x_{i_p})} \right)(x) \neq 0$ *und*

ii) $\det \left(\dfrac{\partial(g_1, \ldots, g_p, f)}{\partial(x_{i_1}, \ldots, x_{i_p}, x_j)} \right)(x) = 0$ *für alle $j = 1, \ldots, n$,*

gilt, so ist x regulär und kritisch für f.

Bemerkung: Die zweite Bedingung enthält nur $(n-p)$ nichttriviale Gleichungen, denn gehört j zum Tupel $I = (i_1, \ldots, i_p)$, so stimmen zwei Spalten der betrachteten Matrix überein, und damit ist deren Determinante immer Null.

Beweis: Wir bezeichnen mit $M^k_{(j_1,\ldots,j_k)}(x)$ den Minor der Jacobi–Matrix $J_{(g,f)}(x)$, der aus den ersten k Zeilen und davon aus den Spalten mit Index $j_1,\ldots,j_k$ gebildet ist. Insbesondere interessieren uns die Minoren zu $k = p$, welche Minoren der Jacobi–Matrix $J_g(x)$ sind, und zu $k = p+1$, welche Minoren der erweiterten Jacobi–Matrix $J_{(g,f)}$ sind. Nach dem Austauschlemma 2.6.2 gilt für diese Minoren und mit Indizes $i_1,\ldots,i_p,\, j_1,\ldots,j_{p+1} \in \{1,\ldots,n\}$

$$M^p_{(i_1,\ldots,i_p)}(x)\, M^{p+1}_{(j_1,\ldots,j_{p+1})}(x) = \sum_{k=1}^{p+1} (-1)^{k+p+1} M^{p+1}_{(i_1,\ldots,i_p,j_k)}(x)\, M^p_{(j_1,\ldots,j_{k-1},j_{k+1},\ldots,j_{p+1})}(x)\,. \quad (2.6)$$

Nach Voraussetzung gilt $M^{p+1}_{(i_1,\ldots,i_p,j)}(x) = 0$ für alle $j = 1,\ldots,n$, daher verschwindet die rechte Seite. Da ebenfalls $M^p_{(i_1,\ldots,i_p)}(x) \neq 0$ vorausgesetzt ist, muss $M^{p+1}_{(j_1,\ldots,j_{p+1})}(x) = 0$ für beliebige Indextupel $(j_1,\ldots,j_{p+1})$ mit $1 \leq j_1 < j_2 < \cdots < j_{p+1} \leq n$ gelten. Damit sind die Anforderungen an einen regulären, kritischen Punkt erfüllt. □

Umgekehrt gibt es für jeden regulären Punkt p Spalten mit Indizes $i_1,\ldots,i_p$ der Jacobi–Matrix $J_g(x)$, welche linear unabhängig sind. Ist x auch kritisch, so verschwindet jeder Minor der Ordnung $p+1$ der Jacobi–Matrix $J_{(g,f)}(x)$, also auch die Minoren $M(i_1,\ldots,i_p,j)(x)$, $j = 1,\ldots,n$.

Wir haben also aus der ursprünglichen Optimierungsaufgabe – nach einer passenden Vertauschung der Variablen – das System

$$x \in \mathbb{R}:\ g_1(x) = \cdots = g_p(x) = 0 = M_{p+1}(x) = \cdots = M_n(x),\ m(x) \neq 0$$

gewonnen, wobei

$$M_k(x) := M^{p+1}_{(1,2,\ldots,p,k)}(x) = \frac{\partial(g_1,\ldots,g_p,f)}{\partial(x_1,\ldots,x_p,x_k)}(x)$$

$$\text{und } m(x) := M^p_{(1,2,\ldots,p)}(x) = \frac{\partial(g_1,\ldots,g_p)}{\partial(x_1,\ldots,x_p)}(x)$$

die „großen“ Minoren der Ordnung $(p+1)$ und der „kleine“ Minor der Ordnung p der Jacobi–Matrix des Systems (g,f) sind. Da nicht bekannt ist, ob der die Regularität der Lösungen sichernde kleine Minor m im Minimalpunkt von Null verschieden ist, müssen alle $\binom{n}{p}$ Möglichkeiten, p der n Spalten auszuwählen und an die ersten p Positionen zu tauschen, durchlaufen werden.

2.6.3 Elemente der Transversalitätstheorie

Für numerische oder numerisch–symbolische Homotopiemethoden zur Lösung von Gleichungssystemen ist es wünschenswert, dass die durch die Funktionen $g_1,\ldots,g_p,\, M_{p+1},\ldots,M_{p+s}$ definierten Hyperflächen sich für jedes $s = 1,\ldots,n-p$ in allgemeiner Lage befinden. Zwei Hyperflächen befinden sich in allgemeiner Lage, wenn sie sich in Schnittpunkten in keiner Richtung berühren, im allgemeineren Fall, wenn die definierenden Funktionen in den gemeinsamen Nullstellen eine Jacobi–Matrix mit vollem Rang besitzen.

Definition 2.6.4 *Seien U, V Untervektorräume von $\mathbb{R}^m$. U wird transversal zu V genannt, wenn ihre Summe schon der gesamte Vektorraum ist, $U + V = \mathbb{R}^m$.*

Seien eine Funktion $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ und ein Untervektorraum E von $\mathbb{R}^m$ gegeben. Φ heißt transversal zu E, wenn für jedes $x \in \mathbb{R}^n$ mit $\Phi(x) \in E$ gilt, dass der Bildraum der Jacobi–Matrix $J_\Phi(x) : \mathbb{R}^n \to \mathbb{R}^m$ transversal zu E ist.

An dieser Definition interessiert uns vor allem der Fall, dass eine Funktion Φ transversal zum Nullunterraum ist. In diesem Fall ist das Urbild $\Phi^{-1}(0)$ nach dem Satz über implizite Funktionen eine differenzierbare Untermannigfaltigkeit des $\mathbb{R}^n$ der Dimension $(n - m)$ oder leer. Ist Φ eine r–fach stetig differenzierbare Funktion, so ist $\Phi^{-1}(0)$ eine C^r–Mannigfaltigkeit.

Einer der Ausgangspunkte der Transversalitätstheorie ist das Lemma von Sard über die Größe der Menge der *kritischen Werte* einer Abbildung. Es sei daran erinnert, dass ein Punkt *kritisch* für eine differenzierbare Abbildung genannt wird, wenn dessen Jacobi–Matrix keinen vollen Rang hat. Ein Wert einer Abbildung heißt dementsprechend kritisch, wenn in seinem Urbild ein kritischer Punkt liegt. Im gegenteiligen Fall heißt der Wert regulär; insbesondere sind Werte – d.h. allgemein Punkte aus dem Wertebereich – deren Urbild leer ist, regulär für eine Abbildung.

Die Menge $W(a, r) := \{x \in \mathbb{R}^m : \max_{k=1,\ldots,m} |x_k - a_k| \leq \frac{r}{2}\}$ wird Würfel um $a \in \mathbb{R}^m$ mit Kantenlänge $r > 0$ genannt. Diesem Würfel wird das Volumen $vol(W(a, r)) := r^m$ zugeordnet.

Definition 2.6.5 *Eine Menge $A \subset \mathbb{R}^m$ wird* vom (m–dimensionalen) Maß Null *genannt, wenn es für jede Schranke $\varepsilon > 0$ eine abzählbare Folge von Würfeln $\{W_k\}_{k\in\mathbb{N}}$, $W_k = W(a_k, r_k)$ gibt, so dass gleichzeitig A von diesen Würfeln überdeckt wird, $A \subset \bigcup_{k\in\mathbb{N}} W_k$, und das gemeinsame Volumen der Würfel durch ε beschränkt ist, d.h. $\sum_{k\in\mathbb{N}} vol(W_k) \leq \varepsilon$.*

Jede abzählbare Vereinigung $\bigcup_{n\in\mathbb{N}} A_n$ von Mengen vom Maß Null ist wieder vom Maß Null. So ist z.B. die Menge der rationalen Zahlen, da abzählbar, eine Menge vom Maß Null.

Das Komplement einer Menge $A \subset \mathbb{R}^m$ vom Maß Null ist dicht in jeder Normtopologie des $\mathbb{R}^n$.

Definition 2.6.6 *Eine offene Menge, deren Komplement vom Maß Null ist, nennt man* residuell dicht, *ebenso jeden abzählbaren Durchschnitt solcher Mengen.*

Nach eben gesagtem ist eine residuell dichte Menge auch im topologischen Sinne dicht, denn das Komplement dieser Menge ist die Vereinigung von abzählbar vielen Mengen vom Maß Null, ist also ebenfalls vom Maß Null.

Satz 2.6.7 (Lemma von Morse–Sard, s. [Hir91, GG86, Dem89]) *Seien $m, n \in \mathbb{N}$, $U \subset \mathbb{R}^n$ eine offene Teilmenge und $\Phi : U \to \mathbb{R}^m$ eine unendlich oft stetig differenzierbare Funktion. Dann ist die Menge der kritischen Werte*

$$C := \left\{ y \in \mathbb{R}^m \middle| \exists x \in U : \Phi(x) = y \;\&\; \operatorname{rank} J_\Phi(x) < \min(m, n) \right\}$$

vom Maß Null. Die Menge der regulären Punkte ist residuell dicht.

Eine Erweiterung des Lemmas von Sard ist der schwache Thomsche Transversalitätssatz. In einer für uns ausreichenden Formulierung lautet er:

Satz 2.6.8 ([Dem89]) *Seien $k, n, m \in \mathbb{N}$ und $U \subset \mathbb{R}^n$ eine offene Teilmenge. Sei $\Phi : \mathbb{R}^k \times U \to \mathbb{R}^m$ eine beliebig oft stetig differenzierbare Abbildung, die transversal zum Nullunterraum von $\mathbb{R}^m$ ist.*

Dann gibt es eine residuell dichte Teilmenge $B \subset \mathbb{R}^k$, so dass für jedes $b \in B$ die Abbildung $\Phi_b : U \to \mathbb{R}^m$, $x \mapsto \Phi_b(x) := \Phi(b, x)$ transversal zum Nullunterraum von $\mathbb{R}^m$ ist.

2.6.4 Optimierungsprobleme in allgemeiner Lage

Definition 2.6.9 *Seien $n \in \mathbb{N}$ und $U \subset \mathbb{R}^n$ eine offene Teilmenge. Eine Folge von differenzierbaren Funktionen $g_1, \ldots, g_n : U \to \mathbb{R}$ heiße transversal, wenn für jedes $k = 1, \ldots, n$ die Funktion $G_k = (g_1, \ldots, g_k) : \mathbb{R}^n \to \mathbb{R}^k$ transversal zum Nullunterraum von $\mathbb{R}^k$ ist.*

Sollen zu einem Optimierungsproblem mit Zielfunktion f und durch Funktionen $g_1, \ldots, g_p$ gegebenen Gleichungsnebenbedingungen alle kritischen Punkte bestimmt werden, so müssen, wie schon angesprochen, alle Varianten, aus n Koordinaten die p auszuwählen, durchlaufen werden.

Definition 2.6.10 *Seien $f, g_1, \ldots, g_p : \mathbb{R}^n \to \mathbb{R}$ mindestens zweimal stetig differenzierbare Funktionen. Das Optimierungsproblem mit Zielfunktion f und Nebenbedingungen $g_1, \ldots, g_p$ werde als* in allgemeiner Lage befindlich *bezeichnet, wenn für jede Permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ die Funktionen $g_1, \ldots, g_p, g_{p+1}, \ldots, g_n$ mit den großen Minoren*

$$g_{p+k} := M^{p+1}_{(\sigma(1),\ldots,\sigma(p),\sigma(p+k))} := \frac{\partial(g_1, \ldots, g_p, f)}{\partial(x_{\sigma(1)}, \ldots, x_{\sigma(p)}, x_{\sigma(p+k)})}$$

eine transversale Folge auf der offenen Teilmenge $U_\sigma := \{x \in \mathbb{R}^n : \mathfrak{m}_\sigma(x) \neq 0\}$ bilden, wobei

$$\mathfrak{m}_\sigma := M^{p}_{(\sigma(1),\ldots,\sigma(p))} := \frac{\partial(g_1, \ldots, g_p)}{\partial(x_{\sigma(1)}, \ldots, x_{\sigma(p)})}$$

der kleine Minor der Permutation ist.

Als Folgerung des Transversalitätssatzes werden wir im Folgenden zeigen, dass eine beliebig kleine lineare Modifikation der Zielfunktion f ausreicht, um ein Optimierungsproblem in allgemeiner Lage zu erhalten. Dazu betrachten wir zunächst jede Permutation einzeln. Dazu ist es o.B.d.A. ausreichend, die identische Permutation zu betrachten.

Satz 2.6.11 (vgl. [BGHM01], Thm. 1) *Seien $n, r \in \mathbb{N}$ mit $r > n$ und seien $g_1, \ldots, g_p \in C^\infty(\mathbb{R}^n)$ beliebig oft stetig differenzierbare Funktionen. Sei*

$$\mathfrak{m} := \det \frac{\partial(g_1, \ldots, g_p)}{\partial(x_1, \ldots, x_p)} \in C^r(\mathbb{R}^n)$$

der kleine Minor zu den ersten p Koordinaten. Mit diesem sei die offene Menge

$$U := \{x \in \mathbb{R}^n : \mathfrak{m}(x) \neq 0\}$$

definiert. Seien weiter für jedes $a \in \mathbb{R}^n$ *die Funktion* $f^a(x) := f(x) + \sum_{k=1}^n a_k x_k$ *sowie die großen Minoren*

$$M^a_{p+k} := \det \frac{\partial(g_1, \ldots, g_p, f^a)}{\partial(x_1, \ldots, x_p, x_{p+k})} \in C^r(\mathbb{R}^n)$$

definiert. Dann gibt es eine residuell dichte Teilmenge $B \subset \mathbb{R}^n$*, so dass für jedes* $b \in B$ *die Funktionen* $g_1, \ldots, g_p, M^b_{p+1}, \ldots, M^b_n : U \to \mathbb{R}$ *eine transversale Folge bilden. Insbesondere ist für jedes* $s = 1, \ldots, n-p$ *die Teilmenge*

$$V_{p+s} := \{x \in \mathbb{R}^n : g_1(x) = \cdots = g_p(x) = 0, \\ M^b_{p+1}(x) = \cdots = M^b_{p+s}(x) = 0, \mathfrak{m}(x) \neq 0\} \subset U$$

eine differenzierbare Mannigfaltigkeit der Kodimension $p + s$*.*

Beweis: Da $\mathfrak{m}$ nach Voraussetzung stetig ist, ist $U := \{x \in \mathbb{R}^n : \mathfrak{m}(x) \neq 0\}$ eine offene Menge. Sei zunächst ein $s \in \{1, \ldots, n-p\}$ fixiert. Die Abbildung

$$\Phi : \mathbb{R}^s \times U \to \mathbb{R}^{p+s}, \quad (a, x) \mapsto (g_1(x), \ldots, g_p(x), M^a_{p+1}(x), \ldots, M^a_{p+s}(x)) \,,$$

hat eine Jacobi–Matrix $J_\Phi(a, x)$, deren Spalten zu den Ableitungen in Richtung $x_1, \ldots, x_p$, $a_{p+1}, \ldots, a_{p+s}$ zusammengefasst die Gestalt

$$\frac{\partial(g_1, \ldots, g_p, M^a_{p+1}, \ldots, M^a_{p+s})}{\partial(x_1, \ldots, x_p, a_{p+1}, \ldots, a_{p+s})} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_p} & & \\ \vdots & & \vdots & 0 & \\ \frac{\partial g_p}{\partial x_1} & \cdots & \frac{\partial g_p}{\partial x_p} & & \\ & & & \mathfrak{m} & 0 \\ & * & & \ddots & \\ & & & 0 & \mathfrak{m} \end{pmatrix}$$

haben. Dabei bezeichnet $*$ den Teil der – nicht weiter interessierenden – partiellen Ableitungen der M^a_{p+k} nach $x_1, \ldots, x_p$. Die partiellen Ableitungen nach $a_{p+1}, \ldots, a_{p+s}$ haben die angegebene Form, denn $g_1, \ldots, g_p$ sind von a unabhängig und es gilt

$$M^a_{p+k} = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_p} & \frac{\partial g_1}{\partial x_{p+k}} \\ \vdots & & \vdots & \vdots \\ \frac{\partial g_p}{\partial x_1} & \cdots & \frac{\partial g_p}{\partial x_p} & \frac{\partial g_p}{\partial x_{p+k}} \\ \frac{\partial f}{\partial x_1} + a_1 & \cdots & \frac{\partial f}{\partial x_p} + a_p & \frac{\partial f}{\partial x_{p+k}} + a_{p+k} \end{pmatrix}$$

und damit für $l = 1, \ldots, s$ nach Entwicklung nach der letzten Zeile

$$\frac{\partial M^a_{p+k}}{\partial a_{p+l}} = \delta_{k,l} \det \frac{\partial(g_1, \ldots, g_p)}{\partial(x_1, \ldots, x_p)} = \delta_{k,l} \mathfrak{m} \,.$$

$\delta_{k,l}$ ist das Kronecker–Symbol, d.h. steht für die Einträge der Einheitsmatrix. Somit hat $J_\Phi(a, x)$ für $x \in U$ mindestens den Rang $p + s$, da dies der Zeilenanzahl entspricht, ist es der maximale Rang.

Nach der angegebenen Version 2.6.8 des Transversalitätssatzes gibt es eine residuell dichte Teilmenge $B_s \subset \mathbb{R}^n$, so dass für jedes $b \in B_s$ die Abbildung Φ_b transversal zum Nullunterraum ist. Dies ist gleichbedeutend dazu, dass $\Phi^{-1}(0) \cap U$ eine differenzierbare Untermannigfaltigkeit der Kodimension $p + s$ ist.

Für die Aussage des Satzes bilden wir den Durchschnitt $B := B_1 \cap \cdots \cap B_{n-p}$, da alle $B_1, \ldots, B_{n-p}$ residuell dicht sind, ist auch B residuell dicht. □

Korollar 2.6.12 *Seien $n \in \mathbb{N}$ und $g_1, \ldots, g_p \in C^\infty(\mathbb{R}^n)$ beliebig oft stetig differenzierbare Funktionen. Dann gibt es eine residuell dichte Teilmenge $B \subset \mathbb{R}^n$, so dass für jedes $b \in B$ das Optimierungsproblem mit Zielfunktion f^b, $x \mapsto f^b(x) := f(x) + b^t x$, und Gleichungsnebenbedingungen $g_1, \ldots, g_p$ in allgemeiner Lage ist.*

Beweis: Seien σ eine Permutation der Indexmenge $\{1, \ldots, n\}$ und $\Pi^\sigma : \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto \Pi^\sigma(x) := (x_{\sigma(1)}, \ldots, x_{\sigma(1)})$ die zugehörige Vertauschung der Koordinaten. Dann kann der vorstehende Satz auf die Funktionen $f^\sigma, g_1^\sigma, \ldots, g_1^\sigma$ mit $f^\sigma(x) := (f \circ \Pi^\sigma)(x) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ usw. angewandt werden. In der daraus folgenden residuell dichten Parametermenge $\tilde{B} \subset \mathbb{R}^n$ kann die Permutation nun zurückgenommen werden, sei $B^\sigma := (\Pi^\sigma)^{-1}(\tilde{B})$.

Es gibt nur endlich viele Permutationen von n Elementen, die Menge $B := \bigcap_{\sigma \in \mathcal{S}^n} B^\sigma$ ist somit ebenfalls residuell dicht. Die mit $b \in B$ gebildeten Optimierungsprobleme mit Zielfunktion f^b und Nebenbedingungen $g_1, \ldots, g_p$ erfüllen nach Konstruktion alle Anforderungen, um sich in allgemeiner Lage zu befinden. □

2.6.5 Polynomiale Optimierungsaufgaben

Die Aussage von Korollar 2.6.12 ist insbesondere dann erfüllt, wenn die das Optimierungsproblem definierenden Funktionen Polynome sind. In diesem Fall kann auch die kritische Menge der ungeeigneten linearen Modifikationen der Zielfunktion als algebraische Varietät dargestellt werden. Diese füllt nicht den gesamten Raum, da ihr Komplement residuell ist. Geht man von Polynomen mit reellen oder rationalen Koeffizienten aus, so ist die kritische algebraische Varietät als Nullstellenmenge von Polynomen mit gleichfalls reellen bzw. rationalen Koeffizienten darstellbar.

Im Umkehrschluss folgt, dass es eine dichte Menge reeller bzw. rationaler linearer Modifikationen der Zielfunktion gibt, so dass das modifizierte Optimierungsproblem sich in allgemeiner Lage befindet.

Satz 2.6.13 *Seien $f, g_1, \ldots, g_p \in \mathbb{Q}[X_1, \ldots, X_n]$ Polynome, so dass $g_1, \ldots, g_p$ eine transversale Folge in $\mathbb{C}^n$ bilden. Dann gibt es eine Teilmenge $B \subset \mathbb{Q}^n$, deren Komplement in einer $\mathbb{Q}$–definierten algebraischen Hyperfläche enthalten ist, und so dass für jeden Parametervektor $a \in B$ und jede Permutation $\sigma \in \mathcal{S}^n$ die Polynome*

$$g_1, \ldots, g_p, M_{p+1}^\sigma, \ldots, M_n^\sigma \in \mathbb{Q}[X_1, \ldots, X_n]$$

eine transversale Folge außerhalb der Hyperfläche $V(m^\sigma)$ bilden. Dabei sind

$$\begin{aligned} m^\sigma :=& M^p_{(\sigma(1),\ldots,\sigma(p))} &=& \frac{\partial(g_1,\ldots,g_p)}{\partial(x_{\sigma(1)},\ldots,x_{\sigma(p)})} \\ M^\sigma_{p+k} :=& M^{p+1}_{(\sigma(1),\ldots,\sigma(p),\sigma(p+k))} &=& \frac{\partial(g_1,\ldots,g_p,f^a)}{\partial(x_{\sigma(1)},\ldots,x_{\sigma(p)},x_{\sigma(p+k)})}, \quad k=1,\ldots,n-p \end{aligned}$$

der „kleine" und die „großen" Minoren der Jacobi–Matrix des Optimierungsproblems zur Permutation σ.

Beweis: Die Eigenschaft eines Vektors $a \in \mathbb{C}^n$, eine ungeeignete lineare Modifikation des Optimierungsproblems zu ergeben, wird durch das Verschwinden von Determinanten von ersten und zweiten partiellen Ableitungen von f^a und $g_1,\ldots,g_p$ auf irreduziblen Komponenten der durch $g_1,\ldots,g_p$ definierten algebraischen Varietäten angegeben. Damit ist die kritische Menge ungeeigneter Modifikationen eine Q–definierbare algebraische Varietät. Ihr Komplement in $\mathbb{C}^n$ ist residuell, damit insbesondere nicht leer.

Daher muss es ein vom Nullpolynom verschiedenes $h \in \mathbb{Q}[Z_1,\ldots,Z_n]$ geben, so dass die kritische Menge in der Hyperfläche $V(h) \subset \mathbb{C}^n$ enthalten ist. Das Komplement $B := \mathbb{Q}^n \setminus V(h)$ dieser Hyperfläche in $\mathbb{Q}^n$ muss dann ebenfalls mindestens einen Punkt enthalten und ist damit in jeder starken Topologie und auch in der $\mathbb{Q}$–Zariski–Topologie dicht. □

Jedes polynomiale Optimierungsproblem in allgemeiner Lage hat nur eine endliche Anzahl regulärer kritischer Punkte. Denn diese sind die isolierten Nullstellen einer endlichen Anzahl polynomialer Systeme, jedes System wiederum hat nur eine endliche Anzahl isolierter Lösungen.

Korollar 2.6.14 *Seien die Polynome $f, g_1,\ldots,g_p \in \mathbb{Q}[X_1,\ldots,X_n]$ durch ein Straight–Line–Programm mit Zeitbedarf T gegeben. Wenn diese Polynome ein polynomiales Optimierungsproblem in allgemeiner Lage bilden, dann kann eine geometrische Lösung der Menge aller regulären kritischen Punkte mittels eines probabilistischen Algorithmus mit Laufzeit*

$$O\left(\tbinom{n}{p} n^6 (T+n^2) M(d\tilde{\delta})^2\right)$$

bestimmt werden.

Beweis: Zu jedem regulären kritischen Punkt $x \in V(g_1,\ldots,g_p)$ gibt es eine Permutation $\sigma \in \mathcal{S}^n$, deren kleiner Minor m^σ in diesem nicht verschwindet. Die großen Minoren dieser Permutation ergeben dann ein transversales Gleichungssystem für x. Dabei kommt es auf die Reihenfolge der ersten p und der weiteren $n-p$ Einträge der Permutation nicht an. Betrachtet man also für jede der $\binom{n}{p}$ Zerlegungen der Menge $\{1,\ldots,n\}$ in Teilmengen zu p und $n-p$ Elementen eine Permutation, so ergeben die zugehörigen Gleichungssysteme alle regulären kritischen Punkte.

Um die Minoren auszuwerten, muss die Jacobi–Matrix bestimmt werden. Der zusätzliche Aufwand dazu beträgt $O(pnT)$. Danach werden $n-p$ große Minoren bestimmt. Im Rahmen

des Berkowitz–Algorithmus kann dies weitgehend parallel erfolgen, der zusätzliche Aufwand zur Auswertung der Minoren hat einen Zeitbedarf von $O(p^4)$. Der Zeitaufwand zur Auswertung des Systems

$$g_1 = \cdots = g_p = M^\sigma_{p+1} = M^\sigma_n = 0, \quad m^\sigma \neq 0$$

läßt sich daher als $\tilde{T} = O(npT + p^4)$ angeben. Der Grad $\tilde{d}$ der Minoren ist durch pd beschränkt, und damit der geometrische Grad des Systems durch $\tilde{\delta} \leq \delta(p\delta)^{n-p}$.

Auf dieses System kann nun die TERA–Kronecker–Methode aus Abschnitt 2.5angewandt werden. Die Laufzeit dieses Algorithmus hat nach Satz 2.5.13 die Abschätzung

$$O\left(n(n\tilde{T} + n^\Omega)M(\tilde{d}\tilde{\delta})^2\right) = O\left(n(n^2pT + n^5)\, p^2 M(d\tilde{\delta})^2\right) = O\left(n^6(T + n^2)M(d\tilde{\delta})^2\right) .$$

□

Jede Permutation $\sigma \in \mathcal{S}^n$ kann als lineare Koordinatentransformation mit Matrix $A = \sum_{k=1}^n \mathbf{e}_{\sigma(k)} \mathbf{e}_k \sigma(k)^t$ aufgefasst werden. Für eine beliebige invertierbare Matrix $A \in \mathbb{Q}^{n\times n}$ kann, analog zur Permutation, das System

$$g_1 = \cdots = g_p = M^A_{p+1} = \cdots = M^A_n = 0, \quad m^A \neq 0$$

aufgestellt werden. Dabei erzeugt A eine Koordinatentransformation $x = Ay$ und

$$\begin{aligned} m^A :=& \frac{\partial(g_1, \ldots, g_p)}{\partial(y_1, \ldots, y_p)} \\ M^A_{p+k} :=& \frac{\partial(g_1, \ldots, g_p, f)}{\partial(y_1, \ldots, y_p, y_{p+k})}, \quad k = 1, \ldots, n-p \end{aligned}$$

sind die Minoren der Ordnung p bzw. $p+1$ zur Jacobi–Matrix mit den partiellen Ableitungen

$$\frac{\partial g_m(Ay)}{\partial y_k} = \sum_{j=1}^n \frac{\partial g_m}{\partial x_j}(x)\, a_{j,k} \quad \text{und} \quad \frac{\partial f(Ay)}{\partial y_k} = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(x)\, a_{j,k}$$

als Komponenten.

Die Menge aller Matrizen A mit rationalen Einträgen, welche invertierbar sind und für die das System $g_1 = \cdots = g_p = M^A_{p+1} = \cdots = M^A_n = 0$, $m^A \neq 0$ transversal ist, ist nichtleer, da in ihr die Permutationsmatrizen enthalten sind. Analog zur Argumentation zur generischen Modifikation der Zielfunktion ist die Menge der geeigneten Matrizen im Komplement einer $\mathbb{Q}$–definierbaren algebraischen Hyperfläche enthalten.

Das Verschwinden des „kleinen“ Minors m^A in einem der regulären kritischen Punkte des Optimierungsproblems kann ebenfalls durch eine $\mathbb{Q}$–definierbare algebraische Hyperfläche im Raum der Matrizen erfasst werden. Eine Matrix A außerhalb der Vereinigung beider Hyperflächen ergibt ein System, dessen Lösungsmenge schon alle regulären kritischen Punkte enthält.

Damit die Betrachtung der in diesem weiteren Sinne generischen Transformation des Optimierungsproblems zur algorithmischen Bestimmung aller regulären kritischen Punkte ausreicht, muss gefordert werden, dass die Folge $g_1, \ldots, g_p, M^A_{p+1}, \ldots, M^A_n$ nicht nur außerhalb der Hyperfläche $V(m^A)$ eine transversale Folge bildet, sondern sogar auf der größeren Menge aller regulären Punkte. Der Nachweis der Existenz derart generischer Koordinatentransformationen ist Inhalt aktueller Arbeiten zu *klassischen* (s [BGHM01]) und *dualen polaren Varietäten* (s. [BGHP05]).

2.6.6 Polare Varietäten

Seien $g_1, \ldots, g_p \in \mathbb{Q}[X_1, \ldots, X_n]$ Polynome, welche eine transversale Folge in $\mathbb{Q}[X_1, \ldots, X_n]$ bilden. Sei $f \in \mathbb{Q}[X_1, \ldots, X_n]$ ein weiteres Polynom, mit welchem sich ein Optimierungsproblem in allgemeiner Lage ergibt. Sei weiter angenommen, dass die identische Koordinatentransformation I die oben angegebenen Eigenschaften hat, d.h. die Folge $g_1, \ldots, g_p, M^I_{p+1}, \ldots, M^I_n$ sei transversal außerhalb $V(m^I)$.

Die Lösungsmengen der Teilsysteme

$$x \in \mathbb{C}^n : \quad g_1(x) = \cdots = g_p(x) = 0, \quad M^I_{p+1}(x) = \cdots = M^I_{p+k}(x), \quad m^I(x) \neq 0$$

können geometrisch interpretiert werden. Sei dazu x irgendein Punkt aus $\mathcal{V} := V(g_1, \ldots, g_p)$, auf dem der kleine Minor $m^I(x)$ nicht verschwindet. Ist nun $k \in 1, \ldots, n-p$ ein Index,für welchen $M^I_{p+k}(x) \neq 0$ gilt, so ist das mit der Jacobi–Matrix gebildete lineare Gleichungssystem

$$u \in \mathbb{C}^n : \quad J_{(g,f)}(x)\, u = \frac{\partial(g_1, \ldots, g_p, f)}{\partial(x_1, \ldots, x_n)}(x)\, u = b$$

für jeden Spaltenvektor b der Länge $(p+1)$ lösbar. Dabei können die Koordinaten $u_{p+1}, \ldots, u_{p+k-1}, u_{p+k+1}, \ldots, u_n$ frei gewählt werden. Gibt man einen beliebigen Spaltenvektor $w \in \mathbb{C}^n$ vor, so ist also auch das mit diesem parametrisierte lineare Gleichungssystem

$$u \in \mathbb{C}^n : J_g(x)\, u = 0, \quad J_f(x)\, u = J_f(x)\, v, \quad e^t_{p+k+1}\, u = e^t_{p+k+1}\, v, \ldots, e^t_n\, u = e^t_n\, v$$

lösbar. Die erste Bedingung besagt, dass u ein Tangentialvektor an $\mathcal{V}$ ist, $u \in T_x\mathcal{V}$. Die weiteren Bedingungen besagen, dass es einen Vektor $v := w - u$ gibt, der zu den Vektoren grad f_x, $e_{p+k+1}, \ldots, e_n$ senkrecht steht. Dabei ist der Gradient von f und die Orthogonalität im Komplexen nicht mit einem hermiteschen Skalarprodukt definiert, sondern mit der komplex–linearen Fortsetzung des euklidischen Skalarprodukts. D.h. es sei eine nicht ausgeartete Bilinearform durch

$$\langle\, a,\, b \,\rangle = \sum_{k=1}^{n} a_n b_n \quad \forall a, b \in \mathbb{C}^n$$

definiert. Dann ergibt sich der Gradient durch Transponieren, grad $f_x = J_f(x)^t$. Sei weiter daran erinnert, dass das orthogonale Komplement – hier der duale Unterraum bzgl. der Bilinearform $\langle\, \cdot,\, \cdot \,\rangle$ – einer Menge $M \subset \mathbb{C}^n$ der Unterraum

$$M^\perp := \{u \in \mathbb{C}^n : \langle\, u,\, v \,\rangle = 0 \; \forall v \in M\}$$

ist. Somit gilt im Punkt x

$$\mathbb{C}^n = T_x\mathcal{V} + \{\operatorname{grad} f_x, e_{p+k+1}, \ldots, e_n\}^\perp .$$

Der Tangentialraum $T_x\mathcal{V}$ ist transversal zum orthogonalen Komplement der angegebenen Vektoren,

$$T_x\mathcal{V} \pitchfork \{\operatorname{grad} f_x, e_{p+k+1}, \ldots, e_n\}^\perp .$$

Ist diese Bedingung verletzt, so kann unter der Voraussetzung $m^I(x) \neq 0$ auf das Verschwinden der großen Minoren im gewählten Punkt x, $M^I_{p+1}(x) = \cdots = M^I_{p+k}(x) = 0$, geschlossen werden, d.h. es gilt $x \in \mathcal{V}_{p+k}$.

Mit der Transversalitätsbedingung ist also eine weitere Charakterisierung der Zwischenvarietäten gewonnen. Diese kann auf natürliche Weise verallgemeinert werden. Seien $a_1, \ldots, a_n \in \mathbb{Q}^n$ eine Basis von $\mathbb{C}^n$. Für jedes $k = 1, \ldots, n$ kann dann eine Menge $\widehat{\mathcal{V}}_k$ als algebraischer Abschluss von

$$\{x \in \mathcal{V} : x \text{ regulär} \quad \& \quad T_x\mathcal{V} \not\pitchfork \{a_{k+1}, \ldots, a_n, \operatorname{grad} f_x\}^\perp\}$$

definiert werden. Dann sind die Varietäten $\widehat{\mathcal{V}}_1, \ldots, \widehat{\mathcal{V}}_p$ aus Dimensionsgründen mit $\mathcal{V}$ identisch, die Vektoren $a_1, \ldots, a_p$ werden also nicht verwendet.

Ist $f \in \mathbb{Q}[X_1, \ldots, X_n]$ linear, so ist $\operatorname{grad} f$ ein konstanter Vektor $b \in \mathbb{Q}^n$. In diesem Fall sei vorausgesetzt, dass die Basisvektoren $a_{k+1}, \ldots, a_n$ so gewählt sind, dass b von ihnen linear unabhängig ist. Das orthogonale Komplement

$$\{a_{k+1}, \ldots, a_n, \operatorname{grad} f_x\}^\perp = \{a_{k+1}, \ldots, a_n, b\}^\perp$$

ist für alle $x \in \mathcal{V}$ derselbe Unterraum der Dimension $k-1$.

Für eine quadratische Zielfunktion der Form $f := \frac{1}{2}\sum_{m=1}^n (X_m - b_m)^2$ ergibt sich $\operatorname{grad} f_x = x - b$. Das orthogonale Komplement

$$\{a_{k+1}, \ldots, a_n, \operatorname{grad} f_x\}^\perp = \{a_{k+1}, \ldots, a_n, x-b\}^\perp$$

ist somit vom Punkt $x \in \mathcal{V}$ abhängig. Für die Punkte aus der affinen Ebene

$$K_k := \{b + \sum_{m=k+1}^n u_m a_m : \quad u_{k+1}, \ldots, u_n \in \mathbb{C}\}$$

sind die Vektoren, zu denen das orthogonale Komplement gebildet wird, linear abhängig. Für alle Punkte außerhalb dieser affinen Ebene ist das orthogonale Komplement ein Unterraum der Dimension $k-1$.

Definition 2.6.15 (vgl. [BGHM01, BGHP05]) *Seien $g_1, \ldots, g_p \in \mathbb{Q}[X_1, \ldots, X_n]$ Polynome, die eine transversale Folge bilden, und sei $\mathcal{V} := V(g_1, \ldots, g_p) \subset \mathbb{C}^n$ deren algebraische Varietät. Seien weiter die Vektoren $b, a_{p+1}, \ldots, a_n \in \mathbb{Q}^n$ linear unabhängig.*

Die klassischen affinen polaren Varietäten *zur Flagge $L_{p+1} \subset \cdots \subset L_n \subset \mathbb{C}^n$ der Unterräume*

$$L_{p+k} := \{a_{k+1}, \ldots, a_n, b\}^\perp$$

sind für $k = 1, \ldots, n-p$ als algebraischer Abschluss

$$\widehat{\mathcal{V}}_{p+k} := \overline{\{x \in \mathcal{V} : x \text{ regulär} \quad \& \quad T_x\mathcal{V} \not\pitchfork L_{p+k}\}}$$

definiert.

Die dualen affinen polaren Varietäten *zur Flagge $K_{p+1} \supset \cdots \supset K_n = \{b\}$ der affinen Unterräume*

$$K_{p+k} := \{b + v : v \in \operatorname{span}(a_{k+p+1}, \ldots, a_n)\}$$

sind für $k = 1, \ldots, n-p$ als algebraischer Abschluss

$$\widehat{\mathcal{V}}_{p+k} := \overline{\{x \in \mathcal{V} \setminus K_{p+k} : x \text{ regulär} \quad \& \quad T_x\mathcal{V} \not\pitchfork \{x - v : \ v \in K_{p+k}\}^{\perp}\}}$$

definiert.

Im Rahmen der projektiven algebraischen Geometrie können beide Begriffe polarer Varietäten als Spezialfälle verallgemeinerter projektiver polarer Varietäten verstanden werden, zu Geometrie und Eigenschaften dieser polaren Varietäten siehe [BGHP05].

Aus den Überlegungen zum Optimierungsproblem in allgemeiner Lage folgt, dass für eine generische Wahl der Vektoren $a_{p+1}, \ldots, a_n, b$ die regulären kritischen Werte des Polynoms $f = \langle b, X \rangle$ bzw. $f = \frac{1}{2} \langle X - b, X - b \rangle$ in allen klassischen bzw. dualen polaren Varietäten enthalten sind. Betrachtet man nun die reellen Anteile der polaren Varietäten, und ist gesichert, dass jede reelle Zusammenhangskomponente mindestens einen regulären kritischen Punkt von f enthält, so enthält auch jede der polaren Varietäten aus jeder reellen Zusammenhangskomponente mindestens einen Punkt.

Ist der reelle Anteil der Varietät $\mathcal{V}$ kompakt, so gibt es für jede lineare Funktion in jeder Zusammenhangskomponente einen Minimalpunkt.

Satz 2.6.16 (s. [BGHP05], Sätze 8 und 10. s. auch [BGHM01]) *Seien $g_1, \ldots, g_p \in \mathbb{Q}[X_1, \ldots, X_n]$ Polynome, welche eine transversale Folge in $\mathbb{Q}[X_1, \ldots, X_n]$ bilden. Sei weiter vorausgesetzt, dass der reelle Teil $\mathcal{V}_{\mathbb{R}} := V(g_1, \ldots, g_p) \cap \mathbb{R}^n$ kompakt ist und nur $(g_1, \ldots, g_p)$–reguläre Punkte enthält.*

Dann gibt es eine residuelle Teilmenge $B \subset (\mathbb{Q}^n)^{n-p+1}$, so dass für jedes Tupel $(a_{p+1}, \ldots, a_n, b) \in B$ für die klassischen affinen polaren Varietäten $\widehat{\mathcal{V}}_{p+k}$, $k = 1, \ldots, n-p$ gilt:

- *$\widehat{\mathcal{V}}_{p+k}$ ist äquidimensional der Dimension $n - p - k$,*
- *$\widehat{\mathcal{V}}_{p+k} \cap \mathbb{R}^n$ ist glatt,*
- *$\widehat{\mathcal{V}}_{p+k} \cap \mathbb{R}^n$ enthält aus jeder Zusammenhangskomponente von $\mathcal{V}_{\mathbb{R}}$ mindestens einen Punkt.*

Trifft die Annahme, dass der reelle Anteil $\mathcal{V}_{\mathbb{R}}$ kompakt ist, nicht zu, so kann es Zusammenhangskomponenten geben, die für lineare Zielfunktionen keine kritischen Punkte aufweisen. Jedoch ist der Abstand zu einem außerhalb der Varietät liegenden Punkt immer nach unten beschränkt, und es gibt in jeder Zusammenhangskomponente mindestens einen Punkt, der das Minimum des Abstands realisiert. Das oben mittels der bilinearen Form gebildete quadratische Polynom $f = \frac{1}{2} \langle X - b, X - b \rangle$ ist, auf den $\mathbb{R}^n$ eingeschränkt, gerade das Quadrat

des euklidischen Abstands. Mit diesem quadratischen Polynom als Zielfunktion wird man auf das Studium dualer polarer Varietäten gelenkt.

Satz 2.6.17 (s. [BGHP05], Sätze 8 und 10) *Seien $g_1, \ldots, g_p \in \mathbb{Q}[X_1, \ldots, X_n]$ Polynome, welche eine transversale Folge in $\mathbb{Q}[X_1, \ldots, X_n]$ bilden. Sei weiter vorausgesetzt, dass der reelle Teil $\mathcal{V}_{\mathbb{R}} := V(g_1, \ldots, g_p) \cap \mathbb{R}^n$ nur $(g_1, \ldots, g_p)$–reguläre Punkte enthält.*

Dann gibt es eine residuelle Teilmenge $B \subset (\mathbb{Q}^n)^{n-p+1}$, so dass für jedes Tupel $(a_{p+1}, \ldots, a_n, b) \in B$ für die dualen affinen polaren Varietäten $\widehat{\mathcal{V}}_{p+k}$, $k = 1, \ldots, n-p$ gilt:

- *$\widehat{\mathcal{V}}_{p+k}$ ist äquidimensional der Dimension $n - p - k$,*
- *$\widehat{\mathcal{V}}_{p+k} \cap \mathbb{R}^n$ ist glatt,*
- *$\widehat{\mathcal{V}}_{p+k} \cap \mathbb{R}^n$ enthält aus jeder Zusammenhangskomponente von $\mathcal{V}_{\mathbb{R}}$ mindestens einen Punkt.*

Analog zu Korollar 2.6.14 gibt es einen effizienten Algorithmus, der zu jeder der polaren Varietäten eine Lifting–Faser bestimmt. Die Aussage von Satz 2.6.17 kann auf beliebige positiv definite Bilinearformen des $\mathbb{R}^n$ und deren komplex bilineare Fortsetzung verallgemeinert werden.

Seien $g_1, \ldots, g_p \in \mathbb{Q}[X_1, \ldots, X_n]$ Polynome, die zusammen mit ihrer reell-allgebraischen Varietät $\mathcal{V}_{\mathbb{R}} := V(g_1, \ldots, g_p) \cap \mathbb{R}^n$ die Voraussetzungen eines der beiden vorangegangenen Sätze erfüllen. Die Konstruktion der polaren Varietäten wird in beiden Fällen durch eine $n \times n$–Matrix A und einen Spaltenvektor b der Dimension n parametrisiert. Fasst man diese Parameter ebenfalls als Variablen auf, so definiert das System

$$M_{p+1}^{A,b} = \cdots = M_n^{A,b} = 0, \quad m^{A,b} \neq 0$$

eine algebraische Varietät $\tilde{\mathcal{V}} \subset \mathbb{C}^{n^2+n} \times \mathbb{C}^n$. Es gibt eine Zariski–offene und damit dichte Teilmenge von $\mathbb{C}^{n^2+n}$, die eine $\mathbb{Q}$–Zariski–offene Teilmenge von $\mathbb{Q}^{n^2+n}$ enthält, so dass für Parameter (A, b) aus dieser Menge die polaren Varietäten einschließlich der letzten, nulldimensionalen, glatt im Sinne der angegebenen Sätze sind. Die Fasern der „großen“ Varietät $\tilde{\mathcal{V}}$ über solcherart generischen Parametern (A, b) bestehen also aus regulären Punkten bzgl. des angegebenen Systems.

Nach dem Implizite-Funktionen-Theorem gibt es eine Umgebung des Parameterpaars (A, b), welche eineindeutig auf je eine Umgebung in $\tilde{\mathcal{V}}$ jedes der Punkte in der Faser über (A, b) abgebildet werden kann. Anders ausgedrückt gibt es eine eineindeutige Zuordnung der Punkte der Faser über (A, b) zu den Punkten jeder benachbarten Faser. Daher ist die Anzahl der Punkte in jeder generischen Faser, d.h. jeder Faser über einem generischen Parameterpaar (A, b), lokal konstant. Da die Menge der nichtgenerischen Parameter in einer Hyperfläche enthalten ist, und diese eine komplexe Kodimension 1 hat, hat die Menge der nichtgenerischen Parameter eine reelle Kodimension von wenigstens 2. Die Menge der generischen Parameter ist daher zusammenhängend, die Anzahl der Punkte in generischen Fasern ist global konstant. Die Anzahl der reellen Punkte in einer generischen Faser ist keine Konstante, jedoch enthält jede generische Faser zu jeder reellen Zusammenhangskomponente von $\mathcal{V}_{\mathbb{R}}$ mindestens einen Punkt.

In einer nichtgenerischen Faser können keine neuen Punkte aus dem Nichts erscheinen. Es ist nur möglich, dass bei der Verfolgung eines Punktes einer benachbarten generischen Faser in die nichtgenerische Faser dieser nach Unendlich divergiert, oder dass der verfolgte reguläre Punkt in der nichtgenerischen Faser singulär wird. Letzteres ist auch der Fall, wenn mehrere Punkte sich auf denselben Punkt der nichtgenerischen Faser zubewegen.

In der praktischen Anwendung der Wavelet–Konstruktion ist eine quadratische Zielfunktion vorgegeben, deren homogener Anteil zweiten Grades positiv definit ist. Durch Berechnung der polaren Varietäten zu zufällig gewählten linearen Modifikationen b und zufällig gewählten Transformationen A kann mit hoher Sicherheit die Anzahl der Punkte zu generischen Parametern bestimmt werden. Stimmt diese mit der Anzahl der Punkte in der Faser zur unmodifizierten Zielfunktion b überein, so kann man zwar nicht davon ausgehen, dass diese Faser generisch ist. Jedoch enthält auch diese Faser in jeder reellen Zusammenhangskomponente mindestens einen regulären kritischen Punkt. In diesem Fall kann man mit hoher Wahrscheinlichkeit annehmen, dass unter den reellen Punkten der Faser die globalen Minimalpunkte enthalten sind.

Kapitel 3

Diskrete Wavelet–Transformation

Wir wollen auf relativ kurzem Wege die polynomialen Gleichungssysteme und Optimierungsaufgaben ableiten, die sich als Bedingungen beim Entwurf von diskreten Wavelet–Transformationen ergeben und die als Anwendung und Testbeispiele der im vorhergehenden Kapitel dargestellten Methoden und Lösungsverfahren dienen sollen. Diese Ableitung läßt sich vollständig im Rahmen der linearen Algebra der *endlichen Zahlenfolgen* oder allgemeiner der *vektorwertigen Folgen* und linearer Abbildungen dieser Folgenräume, die in einem gleich zu konkretisierenden Sinne *verschiebungsinvariant* sind, darstellen. Diese Art der linearen Algebra ist eine der Grundlagen der Theorie der *zeitdiskreten Signalverarbeitung*.

Im Rahmen einer umfassenderen analytischen Theorie der diskreten Wavelet–Transformation bildet die verschiebungsinvariante lineare Algebra der Folgenräume den „nackten“ Rechenkern. Der Sinn einiger Entscheidungen zu wünschenswerten Eigenschaften und Strukturen der nachfolgend betrachteten linearen Abbildungen wird daher vorerst vage bleiben und sich erst in den zwei nachfolgenden Kapiteln zur Abtastung und Multiskalenanalyse aus analytischen Forderungen erschließen.

3.1 Signalalgebra

In der diskreten Signalverarbeitung hat man es meist mit reellwertigen Folgen oder mit Folgen von Tupeln reeller Zahlen zu tun. Beispielsweise entsteht durch Abtasten von Musik oder Geräuschen eine Folge reeller Zahlen, wenn die Aufnahme Mono ist, und eine Folge von Paaren reeller Zahlen, wenn die Aufnahme Stereo ist. Da auch $\mathbb{R}$ ein $\mathbb{R}$–Vektorraum ist, werden wir alle Folgen als vektorwertig ansehen, vom praktischen Standpunkt aus handelt es sich meist um Spaltenvektoren.

Zunächst schränken wir uns auf endliche Folgen und auf Transformationen ein, die wieder endliche Folgen ergeben. Die Räume endlicher Folgen sind wieder Vektorräume, wenn die Glieder der Folgen selbst zahl– oder vektorwertig sind. Auf diesen kann man also allgemein lineare Abbildungen definieren, auf Eindeutigkeit oder Umkehrbarkeit untersuchen etc.

Die diskrete Signalverarbeitung schränkt die betrachteten linearen Abbildungen auf die Klasse der verschiebungsinvarianten Abbildungen ein. Zentral für den Begriff der Verschiebungsinvarianz sind die Operation der *Indexverschiebung* und weitere lineare Operationen auf dem

Folgenräumen, die gegenüber dieser Verschiebung invariant sind. Solche Operatoren werden auch *Filter* bzw. *Filterbank* genannt. Die Verknüpfung von Filterbänken ist wieder eine Filterbank, weshalb man manchmal dieses Teilgebiet der linearen Algebra in Analogie zur Matrixalgebra als *Signalalgebra* bezeichnet.

Ein Beispiel einer solchen verschiebungsinvarianten Abbildung, die eine endliche Folge $x = \{x_n\}_{n\in\mathbb{Z}}$ in eine endliche Folge $y = \{y_n\}_{n\in\mathbb{Z}}$ überführt, ist gliedweise durch

$$y_n = \sum_{k\in\mathbb{Z}} f_{nq-pk}\, x_k$$

mit $p, q \in \mathbb{N}_{>0}$ und einer endlichen Koeffizientenfolge $f = \{f_n\}_{n\in\mathbb{Z}}$ gegeben. Eine Verschiebung von x um q Glieder ergibt eine Verschiebung von y um p Glieder. Die Folgen x und y, und damit auch f, können reell– oder komplexwertig sein. Es ist aber auch möglich, dass x wie y Folgen von Spaltenvektoren sind, f ist dann eine Folge von Matrizen. Dies verallgemeinernd können für x wie y auch vektorwertige Folgen betrachtet werden, die Glieder der Folge f sind dann lineare Abbildungen.

Nicht jede Transformation mit dieser Art der Verschiebungsinvarianz läßt sich in dieser einfachen Form darstellen. Jedoch kann jede verschiebungsinvariante lineare Abbildung von Räumen endlicher Folgen auf eine endliche Folge linearer Abbildungen der Vektorräume der Folgenglieder zurückführen.

3.1.1 Vektorwertige Folgen

Es wurde bereits angesprochen, dass es sinnvoll ist, endliche Folgen über reellen Vektorräumen als Modell eines zeitdiskreten Signals zu betrachten. In der Fourier–Analyse werden die Glieder dieser reellen Folgen jedoch mit komplexen Koeffizienten multipliziert, so dass es sinnvoll ist, von vornherein mit Folgen über komplexen Vektorräumen zu rechnen.

Dies ist keine große Einschränkung, da aus jedem $\mathbb{R}$–Vektorraum W dessen *Komplexifizierung*, der $\mathbb{C}$–Vektorraum $W^{\mathbb{C}}$, konstruiert werden kann. Dessen Vektoren sind Paare des kartesischen Produkts $W \times W$, von denen die erste Komponente als Realteil und die zweite Komponente als Imaginärteil aufgefasst wird. Die Multiplikation mit einer komplexen Zahl $c = a + ib$ erfolgt dann nach den üblichen Regeln der komplexen Multiplikation, mit $\mathbf{w} = (\mathbf{u}, \mathbf{v}) \in W \times W$ sei

$$c\,\mathbf{w} := (a\,\mathbf{u} - b\,\mathbf{v}, a\,\mathbf{v} + b\,\mathbf{u})\,.$$

Der reelle Vektorraum W ist in $W^{\mathbb{C}}$ in Form der rein reellen Vektoren aus $W \times \{0\}$ enthalten. Mit dieser Vereinbarung kann jedes Element von $\Omega^{\mathbb{C}}$ als $(\mathbf{u}, \mathbf{v}) = \mathbf{u} + i\,\mathbf{v}$ geschrieben werden.

Sei V ein $\mathbb{C}$–Vektorraum. Wir betrachten den Raum $\ell_{\text{fin}}(V) := \ell_{\text{fin}}(\mathbb{Z}, V)$ aller *endlichen vektorwertigen Folgen* $\{a_n\}_{n\in\mathbb{Z}}$ mit Indizes in den ganzen Zahlen, die in fast allen Gliedern den Wert Null annehmen. Eine n–gliedrige bzw. n–elementige Folge habe genau $n \in \mathbb{N}$ von Null verschiedene Glieder, eine Folge der Länge n habe ihre von Null verschiedenen Glieder in einem zusammenhängenden Segment von $n \in \mathbb{N}$ Elementen. Jede endliche Folge kann als Summe von eingliedrigen Folgen dargestellt werden.

Es ist nützlich, eine Methode zu vereinbaren, aus einem Vektor eine vektorwertige Folge zu konstruieren. Dazu definieren wir eine „Multiplikation“, die einem Paar aus $V \times \ell_{\text{fin}}(\mathbb{C})$ eine Folge in $\ell_{\text{fin}}(V)$ zuordnet. Sei $\mathbf{v} \in V$ ein Vektor und $a \in \ell_{\text{fin}}(\mathbb{C})$ eine komplexwertige Folge. Dann sei deren Produkt definiert als

$$va := \{a_n \mathbf{v}\} \in \ell_{\text{fin}}(V) .$$

Der Raum der endlichen Folgen mit Werten in $\mathbb{C}$ hat eine *kanonische Basis*, die aus den einelementigen Folgen $\delta^k = \{\delta^k_n\}_{n\in\mathbb{Z}} \in \ell_{\text{fin}}(\mathbb{C})$, $k \in \mathbb{Z}$, besteht. Diese Folgen besitzen an der Stelle k das Glied 1, alle anderen Glieder sind Null. D.h. es gilt $\delta^k_n = \delta_{k,n}$ mit dem Kronecker–Symbol $\delta_{k,n}$. Diese Folgen bilden die Verallgemeinerung der *kanonischen Basisvektoren* in einem Spaltenvektorraum $\mathbb{C}^n$ auf den unendlichdimensionalen Fall. Damit kann jede endliche Folge $b \in \ell_{\text{fin}}(\mathbb{Z}, V)$ als endliche Summe einelementiger Folgen

$$b = \sum_{n\in\mathbb{Z}:b_n\neq 0} b_n \delta^n$$

geschrieben werden.

Sind U und V zwei $\mathbb{C}$–Vektorräume, so bilden die *linearen Abbildungen* von U nach V wieder einen $\mathbb{C}$–Vektorraum. Dieser wird als $\text{Hom}(U, V)$ notiert, als Raum der *linearen Homomorphismen*. Sind U und V endlichdimensional, so auch $\text{Hom}(U, V)$. Wir werden auch endliche Folgen linearer Operatoren betrachten, diese gehören dem Raum $\ell_{\text{fin}}\big(\text{Hom}(U, V)\big)$ an.

Die *direkte Summe* $V \oplus W$ zweier Vektorräume V und W ist deren kartesisches Produkt $V \times W$, das zum Vektorraum wird, indem die Addition und skalare Multiplikation gliedweise definiert werden. Die Vektoren $\mathbf{v} \oplus \mathbf{w} \in V \oplus W$ seien als zweikomponentige Spaltenvektoren aufgefasst, deren erste Komponente $\mathbf{v}$ und die zweite $\mathbf{w}$ ist. Ist U ein weiterer Vektorraum, so kann eine Abbildung $g : U \to V \oplus W$ in Teilabbildungen $g_V : U \to V$ und $g_W : U \to W$ zerlegt werden, die jeweils die erste bzw. zweite Komponente der direkten Summe ergeben. Im Sinne der Matrixalgebra können die Teilabbildungen zu einem Spaltenvektor zusammengefasst werden. Umgekehrt kann eine Abbildung $f : V \oplus W \to U$ mit einem Zeilenvektor (f_V, f_W) identifiziert werden, so dass

$$f(\mathbf{v} \oplus \mathbf{w}) = (f_V, f_W)\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix} = f_V(\mathbf{v}) + f_W(\mathbf{w})$$

gilt.

Man kann den Folgenraum $\ell_{\text{fin}}(V \oplus W)$ mit der direkten Summe $\ell_{\text{fin}}(V) \oplus \ell_{\text{fin}}(W)$ identifizieren. Dabei wird die direkte Summe von Folgen als binäre Operation $\oplus : \ell_{\text{fin}}(V) \times \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(V \oplus W)$ aufgefasst. Dem Paar der zwei Folgen $a = \{a_n\}_{n\in\mathbb{Z}} \in \ell_{\text{fin}}(V)$ und $b = \{b_n\}_{n\in\mathbb{Z}} \in \ell_{\text{fin}}(W)$ wird dabei die Folge direkter Summen $\{a_n \oplus b_n\}_{n\in\mathbb{Z}} \in \ell_{\text{fin}}(V \oplus W)$ zugeordnet.

Dieser Identität folgend kann jede Abbildung $G : \ell_{\text{fin}}(U) \to \ell_{\text{fin}}(V \oplus W)$ als Spaltenvektor $(G_V, G_W)^t$ der zwei partiellen Abbildungen $G_V : \ell_{\text{fin}}(U) \to \ell_{\text{fin}}(V)$ und $G_W : \ell_{\text{fin}}(U) \to \ell_{\text{fin}}(W)$ dargestellt werden. Eine Abbildung $F : \ell_{\text{fin}}(V \oplus W) \to \ell_{\text{fin}}(U)$ wird analog dazu mit einem Zeilenvektor (F_V, F_W) identifiziert.

Die n–fache direkte Summe eines Vektorraumes V mit sich selbst sei als V^n notiert. Die Vektoren dieser Summe sind Spaltenvektoren der Länge n, deren Komponenten Vektoren aus V sind. Im obigen Sinne sei der Raum $\ell_{\text{fin}}(V^n)$ mit dem Raum $\ell_{\text{fin}}(V)^n$ der Spaltenvektoren der Länge n mit Komponenten aus $\ell_{\text{fin}}(V)$ identifiziert.

3.1.2 Periodische Operatoren

Die einfachsten Operationen auf einem Folgenraum sind die Indexverschiebungen. Diese sind, für jeden Vektorraum V und jedes $k \in \mathbb{Z}$, als lineare Abbildungen $\mathcal{T}_k : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(V)$,

$$a = \{a_n\}_{n\in\mathbb{Z}} \mapsto \mathcal{T}_k(a) := \{a_{n-k}\}_{n\in\mathbb{Z}} ,$$

definiert. Z.B. verschiebt $\mathcal{T} := \mathcal{T}_1$ die gesamte Folge um ein Glied in Richtung wachsender Indizes. Die Verknüpfung zweier Verschiebungen ist wieder eine Verschiebung, $\mathcal{T}_k \circ \mathcal{T}_m = \mathcal{T}_{k+m}$. Insbesondere gelten $\mathcal{T}_k \circ \mathcal{T}_{-k} = \mathcal{T}_0 = id_{\ell_{\text{fin}}(V)}$ und $\mathcal{T} \circ \mathcal{T}_k = \mathcal{T}_{k+1}$, somit ist $\mathcal{T}_k = \mathcal{T}^k$, die Verschiebung um ein $k \in \mathbb{N}$. Diese ist also dasselbe, wie die k–fache Wiederholung der einfachen Verschiebung.

Die k–fache Verschiebung einer eingliedrigen Folge δ^n ist die eingliedrige Folge $\mathcal{T}^k\delta^n = \delta^{k+n}$. Damit kann jede endliche Folge $b \in \ell_{\text{fin}}(V)$ als endliche Summe einelementiger Folgen

$$b = \sum_{n\in\mathbb{Z}:b_n\neq 0} b_n\mathcal{T}^n\ \delta^0$$

geschrieben werden. Für $f \in \ell_{\text{fin}}\big(\text{Hom}(V,W)\big)$ und $\mathbf{v} \in V$ kann das Bild unter f definiert werden, es sei

$$f(\mathbf{v}) := \{f_n(\mathbf{v})\}_{n\in\mathbb{Z}} = \left(\sum_{n\in\mathbb{Z}} f_n\mathcal{T}^n\right)(\mathbf{v}\delta^0) .$$

Damit ist eine lineare Abbildung für einelementige Folgen aus $\ell_{\text{fin}}(V)$ definiert. Diese Operation kann auf natürliche Art auf beliebige endliche Folgen $a \in \ell_{\text{fin}}(V)$ erweitert werden. Das Ergebnis dieser erweiterten Abbildung ist das *Faltungsprodukt* $f(\mathcal{T})(a) := f * a$ einer Folge linearer Abbildungen mit einer Folge von Vektoren,

$$f * a = \left(\sum_{n\in\mathbb{Z}} f_n\mathcal{T}^n\right)\left(\sum_{m\in\mathbb{Z}} a_m\mathcal{T}^m\right)\delta^0 = \sum_{n\in\mathbb{Z}}\left(\sum_{m\in\mathbb{Z}} f_m(a_{n-m})\right)\delta^n .$$

Definition 3.1.1 *Für eine beliebige Folge $f \in \ell_{\text{fin}}\big(\text{Hom}(V,W)\big)$ sei $f(\mathcal{T}) : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ der als* Differenzenoperator *bzw.* Faltungsoperator *bezeichnete lineare Operator*

$$f(\mathcal{T})(a) := \sum_{n\in\mathbb{Z}}\left(\sum_{m\in\mathbb{Z}} f_m(a_{n-m})\right)\delta^n .$$

Nach dieser Konstruktion gilt für beliebige $k \in \mathbb{Z}$ die Identität $f(\mathcal{T})(\mathcal{T}^k a) = \mathcal{T}^k(f(\mathcal{T})(a))$.

Definition 3.1.2 *Seien V, W $\mathbb{C}$–Vektorräume, $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ eine lineare Abbildung sowie $p, q \in \mathbb{N}_{>0}$. F wird* (p,q)–periodisch *genannt, wenn für jedes $k \in \mathbb{Z}$ die Identität*

$$\mathcal{T}^{kp} \circ F = F \circ \mathcal{T}^{kq}$$

besteht.

Bemerkung: In der Literatur zur digitalen Datenverarbeitung findet man auch die Bezeichnung *LTI–Systeme* für Differenzenoperatoren (wobei LTI für engl. „linear time–invariant" steht). Oft, gerade in älterer Literatur, findet sich auch der Begriff *Filter* für $(1,1)$–periodische und *Filterbank* für allgemein (p,q)–periodische lineare Abbildungen.

Ist $a = \{a_n\}_{n\in\mathbb{Z}} \in \ell_{\text{fin}}(V)$ eine endliche Folge und $b = \{b_n\}_{n\in\mathbb{Z}} := F(a) \in \ell_{\text{fin}}(W)$ ihr Bild, so ist das Bild der um kq Glieder verschobenen Folge $\mathcal{T}^{kq}a$ die um kp Glieder verschobene Folge $\mathcal{T}^{kp}b$. Es genügt, in der Definition das Bestehen der einzelnen Identität $\mathcal{T}^p \circ F = F \circ \mathcal{T}^q$ zu verlangen.

Trivialerweise sind alle Verschiebungen $\mathcal{T}^k$, $k \in \mathbb{Z}$, und alle anderen Differenzenoperatoren $(1,1)$–periodisch. Sind zwei Abbildungen derselben Folgenräume (p,q)–periodisch, so auch ihre Summe. Man überlegt sich leicht, dass die Verknüpfung eines (p,q)–periodischen Operators $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ und eines (q,r)–periodischen Operators $G : \ell_{\text{fin}}(U) \to \ell_{\text{fin}}(V)$ einen (p,r)–periodischen Operator $F \circ G$ ergibt. Denn

$$\mathcal{T}^p \circ (F \circ G) = F \circ \mathcal{T}^q \circ G = (F \circ G) \circ \mathcal{T}^r \ .$$

Jede (p,q)–periodische lineare Abbildung ist für jedes $n \in \mathbb{N}_{>0}$ auch (np,nq)–periodisch. Jedoch kann man umgekehrt nicht in jedem Fall gemeinsame Faktoren von p und q kürzen. Als Beispiel diene die $(2,2)$–periodische Abbildung, welche einer Folge $a = \{a_n\}$ die Folge $b = \{b_n\}$ mit $b_{2n} = b_{2n+1} := a_{2n}$ zuordnet. Wird die Folge a um ein Glied verschoben, so wird die Folge b aus den Gliedern mit ungeradem Index anstelle derjenigen mit geradem Index gebildet. Diese Abbildung kann also keinesfalls $(1,1)$–periodisch sein.

3.1.3 Zerlegung in elementare periodische Operatoren

Seien V und W Vektorräume über $\mathbb{C}$ und $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(V)$ ein (p,q)–periodischer linearer Operator. Durch Addition eines Vielfachen von q zu einer gegebenen ganzen Zahl $n \in \mathbb{Z}$ kann man immer einen Rest im Segment $\{0, \ldots, q-1\}$ erhalten. Die Division durch q mit einem solchen Rest ist eindeutig, daher kann eine endliche Folge auch als

$$a = \sum_{n\in\mathbb{Z}} a_n \delta^n = \sum_{r=0}^{q-1} \sum_{k\in\mathbb{Z}} a_{qk+r} \delta^{qk+r} = \sum_{r=0}^{q-1} \sum_{k\in\mathbb{Z}} \mathcal{T}^{qk}(a_{qk+r}\delta^r)$$

dargestellt werden. Wendet man auf diese Darstellung den Operator F an, so ergibt sich

$$F(a) = \sum_{r=0}^{q-1} \sum_{k\in\mathbb{Z}} (F \circ \mathcal{T}^{qk})(a_{qk+r}\delta^r) = \sum_{r=0}^{q-1} \sum_{k\in\mathbb{Z}} (\mathcal{T}^{pk} \circ F)(a_{qk+r}\delta^r) \ .$$

Es reicht also, die Bilder von F für einelementige Folgen mit Träger in $\{0,\dots,q-1\}$ zu kennen. Für jedes $r = 0,\dots,q-1$ und jedes $v \in V$ sind die Glieder der Folge $F(v\delta^r)$ Bilder linearer Abbildungen von V nach W. Der Index dieser Glieder kann wieder nach ihrem Rest unter Division durch p unterschieden werden, seien für $s = 0,\dots,p-1$, $r = 0,\dots,q-1$ sowie $k \in \mathbb{Z}$ die linearen Abbildungen $f_{(s,r),k} : V \to W$ so definiert, dass für jedes $v \in V$

$$F(v\delta^r) = \sum_{s=0}^{p-1} \sum_{k\in\mathbb{Z}} f_{(s,r),k}(v)\delta^{kq+s}$$

gilt. Für beliebige endliche Folgen $a \in \ell_{\text{fin}}(V)$ gilt dann

$$F(a) = \sum_{s=0}^{p-1}\sum_{r=0}^{q-1}\sum_{k,l\in\mathbb{Z}} \mathcal{T}^{kq}\big(f_{(s,r),l}(a_{qk+r})\,\delta^{kq+s}\big) = \sum_{s=0}^{p-1}\mathcal{T}^{s}\sum_{r=0}^{q-1}\sum_{k,l\in\mathbb{Z}} f_{(s,r),l}(a_{q(k-l)+r})\,\delta^{kq+s}.$$

Es ergibt sich also im Index l die Struktur eines Faltungsprodukts der Folge $\{f_{(s,r),k}\}_{k\in\mathbb{Z}}$ linearer Abbildungen mit der Folge $\{a_{r+kq}\}_{k\in\mathbb{Z}}$ von Vektoren. Die Summation über den Index n kann als Teil eines gleichzeitig mit der Faltung ausgeführten Matrix–Vektor–Produkts interpretiert werden, die Summation über m erzeugt aus mehreren separaten Folgen $\{b_{m+kp}\}_{k\in\mathbb{Z}}$, $k = 0,\dots,p-1$ eine einzige Folge b. Fasst man die Verknüpfung von linearer Abbildung und Vektor als eine Art der Multiplikation auf, so kann obige Summe in der folgenden Form geschrieben werden:

$$F(a) = \big(1,\mathcal{T},\dots,\mathcal{T}^{p-1}\big)\sum_{k,l\in\mathbb{Z}}\begin{pmatrix} f_{(0,0),l} & f_{(0,1),l} & \cdots & f_{(0,q-1),l}\\ f_{(1,0),l} & f_{(1,1),l} & \cdots & f_{(1,q-1),l}\\ \vdots & \vdots & & \vdots\\ f_{(p-1,0),l} & f_{(p-1,1),l} & \cdots & f_{(p-1,q-1),l}\end{pmatrix}\begin{pmatrix} a_{0+(k-l)q}\\ a_{1+(k-l)q}\\ \vdots\\ a_{q-1+(k-l)q}\end{pmatrix}\delta^{kp}\,. \tag{3.1}$$

Diese Darstellung des linearen Operators F kann in drei Teilschritte unterteilt werden. Der erste ist das Einteilen der Folge a in Blöcke der Länge q und das Aufstellen einer Folge von Spaltenvektoren, wobei jeder Block einen Spaltenvektor ergibt.

Definition 3.1.3 *Der $(1,q)$–periodische Operator $\big(\downarrow\!\downarrow\mathrm{q}\big) : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(V^n)$,*

$$a \mapsto \big(\downarrow\!\downarrow\mathrm{q}\big)(a) := \sum_{k\in\mathbb{Z}}\begin{pmatrix} a_{0+kq}\\ a_{1+kq}\\ \vdots\\ a_{q-1+kq}\end{pmatrix}\delta^k$$

wird Polyphasenzerlegung *genannt.*

Der zweite Schritt ist die Faltung der Folge $\big(\downarrow\!\downarrow\mathrm{q}\big)(a)$ von Spaltenvektoren mit einer Folge $f \in \ell_{\text{fin}}\big(\mathrm{Hom}(V^q,W^p)\big)$ von Matrizen, deren Komponenten lineare Abbildungen sind.

Definition 3.1.4 *Sei $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ ein (p,q)–periodischer Operator und seien $f_{(m,n),k} : V \to W$ die oben konstruierten Abbildungen. Gibt es nur endlich viele von Null verschiedene $f_{(m,n),k}$ und ist $f \in \ell_{\text{fin}}\big(\mathrm{Hom}(V^q,W^p)\big)$ die Folge der Matrizen*

$$f_k := \begin{pmatrix} f_{(0,0),k} & \cdots & f_{(0,q-1),k}\\ \vdots & & \vdots\\ f_{(p-1,0),k} & \cdots & f_{(p-1,q-1),k}\end{pmatrix}, \qquad k \in \mathbb{Z}\,,$$

so wird der Differenzenoperator $F_{poly} := f(\mathcal{T}) : \ell_{\text{fin}}(V^q) \to \ell_{\text{fin}}(W^q)$ *als* Polyphasenmatrix *von* F *bezeichnet.*

Wie oben angemerkt ergibt die Anwendung von Polyphasenzerlegung und Polyphasenmatrix auf eine Folge $a \in \ell_{\text{fin}}(V)$ die Polyphasenzerlegung $\left(\downdownarrows \mathrm{p}\right) b$ der Ergebnisfolge $b := F(a) \in \ell_{\text{fin}}(W)$. Diese Zerlegung ist rückgängig zu machen.

Definition 3.1.5 *Der* $(p,1)$*–periodische lineare Operator* $\left(\uparrow\uparrow \mathrm{p}\right) : \ell_{\text{fin}}(W^p) \to \ell_{\text{fin}}(W)$*,*

$$a = \sum_{k\in\mathbb{Z}} \begin{pmatrix} a_{1,k} \\ a_{1,k} \\ \vdots \\ a_{p,k} \end{pmatrix} \delta^k \mapsto \left(\uparrow\uparrow \mathrm{p}\right)(a) := \sum_{k\in\mathbb{Z}} \sum_{m:=0}^{p-1} a_{m-1,k}\delta^{m+kp}$$

wird Polyphasenrekonstruktion *genannt.*

Insgesamt erhalten wir $F = \left(\uparrow\uparrow \mathrm{p}\right) \circ F_{poly} \circ \left(\downdownarrows \mathrm{q}\right)$.

Lemma 3.1.6 *Für jeden Vektorraum* V *und jedes* $q \in \mathbb{N}_{>0}$ *sind die Operatoren* $\left(\downdownarrows \mathrm{q}\right)$ *und* $\left(\uparrow\uparrow \mathrm{q}\right)$ *zueinander invers.*

Beweis: Die Polyphasenzerlegung basiert auf der Division mit Rest zum Divisor q. Dem Index $n \in \mathbb{Z}$ der Ausgangsfolge werden der Index k und der Komponentenindex $m+1$ in der Folge der Spaltenvektoren zugeordnet, für die $n = m + kq$ und $m \in \{0, \ldots, q-1\}$ gelten. In der Polyphasenrekonstruktion wird auf die gleiche Weise aus Folgenindex und Komponentenindex der Index der rekonstruierten Folge bestimmt. Da die Division mit Rest in der angegebenen Form immer ausführbar und eindeutig ist, sind die Operationen zueinander invers. □

Die Polyphasenmatrix eines (p,q)–periodischen Operators F ergibt sich also auch als $F_{poly} = \left(\downdownarrows \mathrm{p}\right) \circ F \circ \left(\uparrow\uparrow \mathrm{q}\right)$. Auf diese Weise kann F_{poly} auch konstruiert werden, wenn dieser Operator sich nicht als Differenzenoperator einer endlichen Folge linearer Abbildungen darstellen läßt.

Satz 3.1.7 *Seien* V, W *endlichdimensionale* $\mathbb{C}$*–Vektorräume, sowie* $p, q \in \mathbb{N}_{>0}$*. Zu jedem* (p,q)*–periodischen linearen Operator* $F : \ell(V) \to \ell(W)$ *gibt es eine endliche Folge* $f \in \ell_{\text{fin}}\Big(\mathrm{Hom}\big(V^q, W^p\big)\Big)$*, deren Differenzenoperator* $f(\mathcal{T}) := \sum_{k\in\mathbb{Z}} \mathcal{T}^k f_k$ *die Polyphasenmatrix des Operator* F *ist,*

$$F = \left(\uparrow\uparrow \mathrm{p}\right) \circ f(\mathcal{T}) \circ \left(\downdownarrows \mathrm{q}\right).$$

Beweis: F ist durch die Bilder der eingliedrigen Folgen $v\delta^r$, $v \in V$ und $r = 0, \ldots, q-1$, eindeutig bestimmt. Da V endlichdimensional ist, läßt sich jeder Vektor aus einer Basis $\mathbf{e}_1, \ldots, \mathbf{e}_d$ von V linear kombinieren. Der Operator F ist also schon durch die Bildfolgen einer endlichen Anzahl rd von Folgen $e_k\delta^r$, $k = 1, \ldots, d$, $r = 0, \ldots, q-1$ vollständig bestimmt. Die Vereinigung der Träger der Folgen $F(e_k\delta^r)$ ist somit endlich. Sei diese Teilmenge von $\mathbb{Z}$ in einem Segment $\{pM, \ldots, pN-1\}$ enthalten. Weiter seien die linearen Abbildungen $f_{(s,r),k} : V \to W$ wie oben konstruiert. Dann gilt $f_{(s,r),k} = 0$ für $k \geq N$ oder $k < M$. Fasst man die Abbildungen $f_{(s,r),k} = 0$ mit gleichem Index k zu einer Matrix $f_k : V^q \to W^p$ zusammen, so ist die Folge dieser linearen Abbildungen endlich mit Träger im Segment $\{M, \ldots, N-1\}$. Wie schon gesehen, gilt dann

$F = \left(\uparrow\uparrow \mathrm{p}\right) \circ f(\mathcal{T}) \circ \left(\downarrow\downarrow \mathrm{q}\right)$ mit dem Differenzenoperator $f(\mathcal{T})$ der Folge f. □

Da V^q eine direkte Summe ist, kann man die Teiloperatoren zu den Operatoren der Polyphasenzerlegung und –rekonstruktion betrachten. Sei mit $\left(\uparrow \mathrm{q}\right) : \ell_{\mathrm{fin}}(V) \to \ell_{\mathrm{fin}}(V)$ die erste Komponente von $\left(\uparrow\uparrow \mathrm{q}\right)$ und mit $\left(\downarrow \mathrm{q}\right) : \ell_{\mathrm{fin}}(V) \to \ell_{\mathrm{fin}}(V)$ die erste Komponente von $\left(\downarrow\downarrow \mathrm{q}\right)$ bezeichnet. Dann gelten die Identitäten

$$\left(\uparrow\uparrow \mathrm{q}\right) = \Big(\left(\uparrow \mathrm{q}\right), \mathcal{T} \circ \left(\uparrow \mathrm{q}\right), \ldots, \mathcal{T}^{q-1} \circ \left(\uparrow \mathrm{q}\right) \Big)$$

und

$$\left(\downarrow\downarrow \mathrm{q}\right) = \Big(\left(\downarrow \mathrm{q}\right), \left(\downarrow \mathrm{q}\right) \circ \mathcal{T}^{-1}, \ldots, \left(\downarrow \mathrm{q}\right) \circ \mathcal{T}^{1-q} \Big)^t .$$

Denn es ist $\left(\downarrow \mathrm{q}\right)(\mathcal{T}^{-m}(a)) = \sum_{k\in\mathbb{Z}} a_{qk+m}\delta^k$ die m–te Polyphasenteilfolge und $\mathcal{T}^m(\left(\uparrow \mathrm{q}\right)(a)) = \sum_{k\in\mathbb{Z}} a_k \delta^{m+kq}$ die isolierte Rekonstruktionsvorschrift für die m–te Polyphasenteilfolge. Aus der inversen Beziehung der Polyphasenoperatoren ergeben sich die Identitäten

$$id_{\ell_{\mathrm{fin}}(V)} = \left(\uparrow\uparrow \mathrm{q}\right) \circ \left(\downarrow\downarrow \mathrm{q}\right) = \sum_{m=0}^{q-1} \mathcal{T}^m \circ \left(\uparrow \mathrm{q}\right) \circ \left(\downarrow \mathrm{q}\right) \circ \mathcal{T}^{-m}$$

und

$$\delta_{m,n} id_{\ell_{\mathrm{fin}}(V)} = \Big(id_{\ell_{\mathrm{fin}}(V^q)}\Big)_{m,n} = \left(\left(\downarrow\downarrow \mathrm{q}\right) \circ \left(\uparrow\uparrow \mathrm{q}\right)\right)_{m,n} = \left(\downarrow \mathrm{q}\right) \circ \mathcal{T}^{m-n} \circ \left(\uparrow \mathrm{q}\right)$$

für $m, n \in \{0, \ldots, q-1\}$.

Diese Teiloperatoren ermöglichen eine zweite Sichtweise auf die Struktur einer (p,q)–periodischen Abbildung $F : \ell_{\mathrm{fin}}(V) \to \ell_{\mathrm{fin}}(W)$. Seien aus den Komponentenabbildungen $f_{(s,r),k}$ die Differenzenoperatoren $f_{(s,r)}(\mathcal{T}) := \sum_{k\in\mathbb{Z}} f_{(s,r),k}\mathcal{T}^k$ gebildet. Mit den formalen Regeln der Matrixalgebra gilt dann

$$F = \Big(\left(\uparrow \mathrm{p}\right) \mathcal{T} \left(\uparrow \mathrm{p}\right) \cdots \mathcal{T}^{p-1} \left(\uparrow \mathrm{p}\right) \Big) \begin{pmatrix} f_{(0,0)}(\mathcal{T}) & f_{(0,1)}(\mathcal{T}) & \cdots & f_{(0,q-1)}(\mathcal{T}) \\ f_{(1,0)}(\mathcal{T}) & f_{(1,1)}(\mathcal{T}) & \cdots & f_{(1,q-1)}(\mathcal{T}) \\ \vdots & \vdots & & \vdots \\ f_{(p-1,0)}(\mathcal{T}) & f_{(p-1,1)}(\mathcal{T}) & \cdots & f_{(p-1,q-1)}(\mathcal{T}) \end{pmatrix} \begin{pmatrix} \left(\downarrow \mathrm{q}\right) \\ \left(\downarrow \mathrm{q}\right) \mathcal{T}^{-1} \\ \vdots \\ \left(\downarrow \mathrm{q}\right) \mathcal{T}^{1-q} \end{pmatrix} .$$

Bemerkung: Ist, im Rahmen einer Anwendung in der Signalverarbeitung, eine Intervalllänge $T > 0$, z.B. als Länge eines Zeitintervalls, fixiert, und wird jedem Zeitpunkt nT das Folgenglied a_n zugeordnet, so hat die Folge a eine Datenrate von $1/T$.

- Die Folge $\left(\downarrow \mathrm{q}\right) a$ hat, wenn die zeitliche Zuordnung der Glieder beibehalten wird, eine Datenrate von $1/(qT)$. Deshalb wird diese Operation auch als *Untertaktung* (engl. „downsampling“) bezeichnet.
- Behält man in $\left(\uparrow \mathrm{q}\right)(a)$ die zeitliche Anordnung der Glieder von a bei und fügt die hinzukommenden Nullvektoren gleichmäßig ein, so vergrößert sich die Datenrate von $1/T$ zu q/T. Daher wird diese Operation auch als *Übertaktung* (engl. „upsampling“) bezeichnet.

3.1.4 Invertierbarkeit von Filterbänken

Definition 3.1.8 *Seien V, W zwei $\mathbb{C}$–Vektorräume. Eine lineare Abbildung $G : W \to V$ heißt* linksinvers *zu einer linearen Abbildung $F : V \to W$, falls $G \circ F = id_V$ gilt. In dieser Situation ist F auch* rechtsinvers *zu G. Eine lineare Abbildung heißt* invertierbar*, wenn sie sowohl eine links– als auch eine rechtsinverse Abbildung besitzt.*

Für eine invertierbare Abbildung stimmen die links– und die rechtsinverse Abbildung überein und sind eindeutig bestimmt.

Lemma 3.1.9 *Seien V, W endlichdimensionale $\mathbb{C}$–Vektorräume, $f \in \ell_{\text{fin}}(\text{Hom}(V,W))$ und $g \in \ell_{\text{fin}}(\text{Hom}(W,V))$ endliche Folgen linearer Abbildungen und $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ und $G : \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(V)$ die zugehörigen Differenzenoperatoren $F := f(\mathcal{T}) = \sum_{n\in\mathbb{Z}} f_n \mathcal{T}^n$ bzw. $G := g(\mathcal{T}) = \sum_{n\in\mathbb{Z}} g_n \mathcal{T}^n$. G ist genau dann linksinvers zu F, wenn für die Koeffizientenfolgen beider Operatoren die Identitäten*

$$\sum_{k\in\mathbb{Z}} g_k \circ f_{n-k} = \delta_{n,0}\, id_V$$

für jedes $n \in \mathbb{Z}$ gelten.

Beweis: $H := G \circ F$ ist ebenfalls ein Differenzenoperator. Seine Koeffizientenfolge $h \in \ell_{\text{fin}}(\text{Hom}(V,V))$ ergibt sich aus

$$H = h(\mathcal{T}) = \sum_{k\in\mathbb{Z}} g_k \mathcal{T}^k \circ \sum_{m\in\mathbb{Z}} f_m \mathcal{T}^m = \left(\sum_{n\in\mathbb{Z}} g_k \circ f_{n-k} \right) \mathcal{T}^n$$

als Faltungsprodukt der Koeffizientenfolgen von G und F. Da diese Koeffizientenfolgen endlich sind, ist auch die Koeffizientenfolge von H endlich. Vergleicht man die Bilder einelementiger Folgen auf beiden Seiten von $G \circ F = id_{\ell_{\text{fin}}(V)}$, so ergeben sich die Identitäten der Behauptung. Sind umgekehrt diese Identitäten erfüllt, so ist die Verknüpfung von G und F die identische Abbildung der Folgen über V. □

Seien V, W endlichdimensionale $\mathbb{C}$–Vektorräume, $p, q \in \mathbb{N}_{>0}$ und $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ ein (p,q)–periodischer linearer Operator. Besitzt dieser einen inversen linearen Operator $G : \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(V)$, so ist dieser (q,p)–periodisch. Denn aus $\mathcal{T}^p \circ F = F \circ \mathcal{T}^q$ und $F \circ G = id$, $G \circ F = id$ folgt

$$\mathcal{T}^q \circ G = G \circ F \circ \mathcal{T}^q \circ G = G \circ \mathcal{T}^p \circ F \circ G = G \circ \mathcal{T}^p \,.$$

Ist F nur linksinvertierbar, so kann ein linksinverser Operator $G : \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(V)$, der (q,p)–periodisch ist, konstruiert werden. Sei $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Basis in W. Für jede einelementige Folge $\mathbf{w}_k \delta^j$, $j = 0, \ldots, p-1$ gibt es nach Voraussetzung ein Urbild $a_{k,n} \in \ell_{\text{fin}}(V)$. Dementsprechend ist $\mathcal{T}^{Nq} a_{k,n}$ ein Urbild von $\mathcal{T}^{Np} \mathbf{w}_k \delta^n$. Die Zuordnung der Basisfolgen kann zu einer linearen Abbildung $G : \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(V)$ fortgesetzt werden, diese ist zu F linksinvers und (q,p)–periodisch.

3.2 Orthogonale Wavelet–Filterbänke

Sei V ein hermitescher $\mathbb{C}$–Vektorraum. Dann kann auf $\ell_{\text{fin}}(V)$ ebenfalls ein Skalarprodukt definiert werden, indem das Skalarprodukt von V gliedweise angewandt wird. Je zwei Folgen $a, b \in \ell_{\text{fin}}(V)$ wird dabei die komplexe Zahl

$$\langle a,\, b \rangle_{\ell_2(V)} := \sum_{n\in\mathbb{Z}} \langle a_n,\, b_n \rangle_V$$

zugewiesen. Diese Reihe hat nur endlich viele von Null verschiedene Glieder. Man überzeugt sich leicht, dass dies tatsächlich ein hermitesches Skalarprodukt auf $\ell_{\text{fin}}(V)$ ist. Eingliedrige Folgen mit den nichttrivialen Gliedern an verschiedenen Indizes sind mit diesem Skalarprodukt orthogonal.

Ist W ein weiterer $\mathbb{C}$–Vektorraum mit hermiteschem Skalarprodukt, so kann man nach Filterbänken $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ suchen, welche das Skalarprodukt der Folgen erhalten. Lineare Operatoren mit dieser Eigenschaft werden als *semi–unitär* bezeichnet.

Seien zum Beispiel $V = \mathbb{C}$ und der semi–unitäre Operator ein (p,q)–periodischer Operator $F = (\downarrow\mathrm{q})\, f(\mathcal{T})\, (\uparrow\mathrm{p})$, der durch eine endliche Folge $f \in \ell_{\text{fin}}(\mathbb{C})$ und Zahlen $p, q \in \mathbb{N}_{>0}$ gegeben ist. Aus der Erhaltung der Skalarprodukte unter F ergibt sich für jede Folge $c \in \ell_{\text{fin}}(\mathbb{C})$ die Identität

$$\|c\|_{\ell_2}^2 = \|F(c)\|_{\ell_2}^2 = \sum_{n\in\mathbb{Z}} \left| \sum_{k\in\mathbb{Z}} f_{nq-kp} c_k \right|^2 = \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} \left(\sum_{n\in\mathbb{Z}} f_{nq-kp} \bar{f}_{nq-lp} \right) c_k \bar{c}_l \,.$$

Daraus erhält man die Bedingungen

$$\sum_{n\in\mathbb{Z}} f_{nq-kp} \bar{f}_{nq-lp} = \delta_{k,l} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases} .$$

Für $f \in \ell_{\text{fin}}(\mathbb{R})$ sind dies quadratische Gleichungen in den Gliedern von f. Der erste Index kann auf den Bereich $k = 0, \ldots, q-1$ eingeschränkt werden.

In einer weiteren Spezialisierung dieses Beispiels auf einen $(1,2)$–periodischen unitären Operator $F = (f_0 + f_1\mathcal{T} + f_2\mathcal{T}^2 + f_3\mathcal{T}^3)\, (\uparrow 2)$ mit reellen Koeffizienten erhalten wir für die Semi–Unitarität die zwei Bedingungen

$$f_0^2 + f_1^2 + f_2^2 + f_3^2 = 1 \quad \text{und} \quad f_0 f_2 + f_1 f_3 = 0 \,.$$

Durch Kombination beider Bedingungen ergibt sich weiter, dass sowohl das Paar $(f_0 + f_2, f_1 + f_3)$ als auch das Paar $(f_0 - f_2, f_1 - f_3)$ Punkte auf dem Einheitskreis sind. Diese Punkte können jeweils durch einen Winkel angegeben werden. Parametrisiert man den Winkel des ersten Punktes als $\alpha + \beta$ und den des zweiten mit $\alpha - \beta$, so erhält man die folgende Parametrisierung der orthogonalen 4–Tupel:

$$\begin{aligned} f_0 &= \quad \cos\alpha\, \cos\beta, \quad & f_1 &= \sin\alpha\, \cos\beta, \\ f_2 &= -\sin\alpha\, \sin\beta, \quad & f_3 &= \cos\alpha\, \sin\beta \,. \end{aligned}$$

Man kann in dieser Parametrisierung den Polyphasenvektor von f bzw. die Polyphasenmatrix von F faktorisieren, es gilt

$$\begin{pmatrix} f_0 + f_2\mathcal{T} \\ f_1 + f_3\mathcal{T} \end{pmatrix} = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{T} \end{pmatrix} \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}$$

Eine solche Faktorisierung in offensichtlich orthogonale bzw. unitäre Matrizen und einen Vektor der Länge 1 ist für beliebige semi–unitäre Differenzenoperatoren möglich.

3.2.1 Adjungierte Abbildungen und duale Filterbänke

Definition 3.2.1 *Seien V, W hermitesche $\mathbb{C}$–Vektorräume und $F : V \to W$ eine lineare Abbildung.*

F wird semi–unitär *genannt, wenn für beliebige $a \in V$ die Identität $\|F(a)\|_W = \|a\|_V$ gilt.*

Eine Abbildung $G : W \to V$, mit welcher die Identität

$$\langle\, F(a),\, b \,\rangle_W = \langle\, a,\, g(b) \,\rangle_V$$

für beliebige $a \in V$ und $b \in W$ gilt, wird adjungiert *zu F genannt und mit $F^* := G$ bezeichnet.*

Sind sowohl F als auch F^* semi–unitär, so wird F unitär *genannt.*

Es ist also, wenn es eine adjungierte Abbildung F^* gibt, auch wiederum F adjungiert zu F^*. Die adjungierte Abbildung F^* existiert immer, wenn V und W endlichdimensional sind. Allgemeiner kann die Existenz einer adjungierten Abbildung mittels des *Satz von Riesz* nachgewiesen werden, wenn V und W Hilbert–Räume sind.

Ist F semi–unitär und existiert die adjungierte Abbildung F^*, so folgt für beliebige $a, b \in V$ die Identität

$$\langle\, a,\, b \,\rangle_V = \langle\, F(a),\, F(b) \,\rangle_W = \langle\, F^*(F(a)),\, b \,\rangle_V \ .$$

Daher ist F^* linksinvers zu F. Analog folgt für ein unitäres F, dass $F^* = F^{-1}$ die inverse Abbildung zu F ist. Umgekehrt folgt aus der Existenz einer (links–)inversen adjungierten Abbildung F^*, dass F (semi–)unitär ist.

Sind V und W hermitesche $\mathbb{C}$–Vektorräume, so auch die Folgenräume $\ell_{\text{fin}}(V)$ und $\ell_{\text{fin}}(W)$ mit dem oben definierten Skalarprodukt. Wir können nun nach der Struktur adjungierter Abbildungen unter der zusätzlichen Voraussetzung der Verschiebungsinvarianz fragen.

Lemma 3.2.2 *Seien V, W endlichdimensionale $\mathbb{C}$–Vektorräume, $f \in \ell_{\text{fin}}\big(\text{Hom}(V, W)\big)$ eine endliche Folge linearer Abbildungen und $F := f(\mathcal{T}) := f_{-M}\mathcal{T}^{-M} + \cdots + f_M\mathcal{T}^M : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ der dazugehörige Differenzenoperator.*

Sei $f^* := \{f^*_{-k}\} \in \ell_{\text{fin}}\big(\text{Hom}(W, V)\big)$ die Folge der adjungierten Koeffizientenabbildungen nach Umkehr der Reihenfolge. Dann ist

$$F^* := f^*(\mathcal{T}) = f^*_M\mathcal{T}^{-M} + \cdots + f^*_{-M}\mathcal{T}^M : \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(V)$$

zu $F = f(\mathcal{T})$ adjungiert.

Beweis: Seien $a \in \ell_{\text{fin}}(V)$ und $b \in \ell_{\text{fin}}(W)$. Dann ist

$$\begin{aligned}\langle F(a),\, b\rangle_{\ell_{\text{fin}}(W)} &= \sum_{k=-M}^{M} \left\langle \mathcal{T}^k f_k(a),\, b \right\rangle_{\ell_{\text{fin}}(W)} = \sum_{k=-M}^{M} \sum_{n\in\mathbb{Z}} \langle f_k(a_{n-k}),\, b_n\rangle_W \\ &= \sum_{k=-M}^{M} \sum_{n\in\mathbb{Z}} \langle a_n,\, f_k^*(b_{n+k})\rangle_V = \sum_{k=-M}^{M} \left\langle a,\, \mathcal{T}^{-k} f_k^*(b) \right\rangle_{\ell_{\text{fin}}(V)} \\ &= \langle a,\, F^*(b)\rangle_{\ell_{\text{fin}}(V)}\end{aligned}$$

□

Insbesondere hat der Verschiebungsoperator den adjungierten Operator $\mathcal{T}^* = \mathcal{T}^{-1}$. Neben Differenzenoperatoren können auch die adjungierten der weiteren elementaren verschiebungsinvarianten Operatoren betrachtet werden.

Lemma 3.2.3 *Seien V, W hermitesche $\mathbb{C}$–Vektorräume und $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ eine (p,q)–periodische lineare Abbildung.*

i) Existiert die adjungierte Abbildung $F^* : \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(V)$, so ist sie (q,p)–periodisch.

ii) $(\uparrow\mathrm{p}) : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(V)$ ist adjungiert zu $(\downarrow\mathrm{p}) : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(V)$, insbesondere ist $(\uparrow\mathrm{p})$ semi–unitär.

iii) $(\uparrow\uparrow\mathrm{p})$ ist adjungiert zu $(\downarrow\downarrow\mathrm{p})$, beide Operatoren sind unitär.

Beweis: zu i) Ist F (p,q)–periodisch und die F^* adjungierte Abbildung, so gilt $\mathcal{T}^p \circ F = F \circ \mathcal{T}^q$, und weiter

$$F^* \circ (\mathcal{T}^p)^* = (\mathcal{T}^q)^* \circ F^* \implies \mathcal{T}^q \circ F^* = F^* \circ \mathcal{T}^p\,.$$

Somit ist F^* (q,p)–periodisch.

zu ii) Ist $a \in \ell_{\text{fin}}(V)$, so ist $(\downarrow\mathrm{p})\, a = \{a_{pn}\}_{n\in\mathbb{Z}}$. Somit gilt mit einer weiteren Folge $b \in \ell_{\text{fin}}(V)$

$$\langle\, (\downarrow\mathrm{p})\, a,\, b\,\rangle_{\ell_{\text{fin}}(V)} = \sum_{n\in\mathbb{Z}} \langle\, a_{pn},\, b_n\,\rangle_V = \langle\, a,\, (\uparrow\mathrm{p})\, b\,\rangle_{\ell_{\text{fin}}(V)}$$

Da $(\downarrow\mathrm{p}) \circ (\uparrow\mathrm{p})$ die Identität auf $\ell_{\text{fin}}(V)$ ergibt, ist $(\uparrow\mathrm{p})$ semi–unitär.

zu iii) Seien $a = \{(a_{1,n}, \ldots, a_{p,n})^t\} \in \ell_{\text{fin}}(V^p)$ und $b \in \ell_{\text{fin}}(V)$. Dann gilt

$$\langle\, (\uparrow\uparrow\mathrm{p})(a),\, b\,\rangle_{\ell_{\text{fin}}(V)} = \sum_{m\in\mathbb{Z}} \sum_{r=0}^{p-1} \langle\, a_{r,m},\, b_{pm+r}\,\rangle_V = \langle\, a,\, (\downarrow\downarrow\mathrm{p})(b)\,\rangle_{\ell_{\text{fin}}(V^p)}$$

Da beide Operatoren invers zueinander sind, sind beide auch unitäre Operatoren. □

Zum Beispiel definiert ein Vektor $\mathbf{v}$ in einem hermiteschen $\mathbb{C}$–Vektorraum V eine Abbildung $\mathbf{v} : \mathbb{C} \to V$ mittels der skalaren Multiplikation, $x \mapsto \mathbf{v}(x) := x\mathbf{v}$. Der duale Operator dazu bildet einen beliebigen Vektor $\mathbf{x} \in V$ auf das Skalarprodukt $\langle \mathbf{x},\, \mathbf{v}\rangle_V$ ab. Auf Folgen übertragen definiert jede Folge $a \in \ell_{\text{fin}}(V)$ einen Differenzenoperator $a(\mathcal{T}) : \ell_{\text{fin}}(\mathbb{C}) \to \ell_{\text{fin}}(V)$ mit einem dualen Operator $a(\mathcal{T})^* : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(\mathbb{C})$, der eine Folge $c \in \ell_{\text{fin}}(V)$ auf die Folge $\{\sum_{k\in\mathbb{Z}} \langle c_{n-k},\, a_k\rangle_V\}_{n\in\mathbb{Z}}$ abbildet.

3.2.2 Zur Struktur semi–unitärer Filterbänke

Lemma 3.2.4 *Seien V, W zwei endlichdimensionale hermitesche Vektorräume. Eine $(1,1)$–periodische lineare Abbildung $F = f(\mathcal{T}) = f_{-M}\mathcal{T}^{-M} + \cdots + f_M\mathcal{T}^M : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ ist genau dann semi–unitär, wenn*

$$\sum_{k=-M}^{M} f_k^* \circ f_k = id_V \text{ und } \sum_{k=-M}^{M} f_k^* \circ f_{k+m} = 0 \text{ bei } m \neq 0$$

gelten.

Beweis: Diese Identitäten folgen aus der Gestalt des adjungierten Operators nach Lemma 3.2.2 und den Identitäten für einen linksinversen Differenzenoperator nach Lemma 3.1.9. □

Unter denselben Voraussetzungen ist F genau dann unitär, wenn ebenfalls die Beziehungen

$$\sum_{k=-M}^{M} f_k \circ f_k^* = id_V \text{ und } \sum_{k=-M}^{M} f_k \circ f_{k+m}^* = 0 \text{ bei } m \neq 0$$

gelten.

Sind V und W endlichdimensionale hermitesche $\mathbb{C}$–Vektorräume, $f \in \ell_{\text{fin}}\big(\text{Hom}(V,W)\big)$ und $F = f(\mathcal{T}) : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ semi–unitär, so muss die Dimension von W größer sein als die von V. Denn der Operator $g := \sum_{k\in\mathbb{Z}} f_k : V \to W$ ist nach den Bedingungen in Lemma 3.2.4 ebenfalls semi–unitär und daher insbesondere injektiv. Ist F unitär, so gilt auch die andere Ungleichung, d.h. die unitäre Abbildung g vermittelt eine Isomorphie der gleichdimensionalen Vektorräume V und W.

Für (p,q)–periodische Operatoren $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ mit beliebigen $p, q \in \mathbb{N}_{>0}$ kann man diese Eigenschaften auf deren Polyphasenoperator $F_{poly} = \big(\downarrow\downarrow \text{p}\big) \circ F \circ \big(\uparrow\uparrow \text{q}\big)$ zurückführen, F ist genau dann (semi–)unitär, wenn F_{poly} (semi–)unitär ist. Denn die Operatoren $\big(\downarrow\downarrow \text{p}\big)$ und $\big(\uparrow\uparrow \text{q}\big)$ der Polyphasenzerlegung und –rekombination sind für beliebige $p, q \in \mathbb{N}_{>0}$ unitär. Sind $V = \mathbb{C}^m$ und $W = \mathbb{C}^n$ Spaltenvektorräume, so sind die Koeffizientenabbildungen von F_{poly} Matrizen mit mq Spalten und np Zeilen, ist F semi–unitär, so muss $np \geq mq$ gelten, ist F unitär, so sind beide Dimensionen gleich.

Satz 3.2.5 *Seien V ein endlichdimensionaler hermitescher $\mathbb{C}$–Vektorraum, $a \in \ell_{\text{fin}}\big(\text{Hom}(V,V)\big)$ und $s \in \mathbb{N}_{>0}$. Der $(s,1)$–periodische Operator $F := a(\mathcal{T}) \circ \big(\uparrow \text{s}\big)$ ist genau dann semi–unitär, wenn*

$$\sum_{n\in\mathbb{Z}} a_{n+sj}^* a_{n+sk} = \delta_{k,l} id_V \qquad \textit{(Kronecker–Delta)}$$

für beliebige $k, l \in \mathbb{Z}$ gilt.

Beweis: F hat den adjungierten Operator $F^* = \big(\downarrow \text{s}\big) \circ a^*(\mathcal{T})$. Der zusammengesetzte Operator $F^* \circ F = \big(\downarrow \text{s}\big) \circ a^*(\mathcal{T}) \circ a(\mathcal{T}) \circ \big(\uparrow \text{s}\big)$ ist $(1,1)$–periodisch und besitzt die Koeffizientenfolge

$$\left\{ \sum_{n\in\mathbb{Z}} a_{n+sk}^* \circ a_n \right\}_{k\in\mathbb{Z}} \in \ell_{\text{fin}}\big(\text{Hom}(V,V)\big) \, .$$

Damit F semi–unitär ist, müssen alle Glieder dieser Folge bis auf das Glied zum Index Null verschwinden und im Index Null muss die Identität auf V anzufinden sein. □

Im Fall $V = \mathbb{C}$ ergeben sich daraus Beziehungen für die komplexwertige Folge $a \in \ell_{\text{fin}}(\mathbb{C})$,

$$\sum_{n\in\mathbb{Z}} \bar{a}_n a_{n+sm} = \delta_{0,m} \,.$$

Sind die Glieder der Folge a reell, so ist dies ein System quadratischer Gleichungen.

3.3 Symmetrische Filterbänke

Oft bearbeitet man relativ kurze Stücke von – zumindest theoretisch – sehr langen, sich langsam ändernden Folgen. Die Werte außerhalb eines solchen kurzen Stücks stehen nicht zur Verfügung. Wendet man eine endlich definierte Filterbank auf einen solchen endlichen Ausschnitt einer Folge an, so ergeben sich Fehler in einer Umgebung der Enden des Ausschnitts. Diese versucht man durch eine geeignete Fortsetzung an den Rändern zu vermindern, ohne die resultierende Folge zu stark zu verlängern.

Ist nichts über die Natur der Folge bekannt, so ist die beste Schätzung der Folge außerhalb des bekannten Ausschnitts durch die Fortsetzung mit Nullgliedern gegeben. Es entsteht eine endliche Folge, die Sprünge an den Enden des Ausschnitts aufweist. Hat man jedoch eine sich langsam ändernde Folge gegeben, so zerstören die Sprünge an den Enden des Ausschnitts, den Charakter der langsamen Änderung von Folgeglied zu Folgeglied.

Eine einfache Möglichkeit, ein kurzes Stück einer Folge unter Beibehaltung der Charakteristik der Folge fortzusetzen, ist die periodische Wiederholung der Folge. Diese Fortsetzung kann zwar einen ebenso großen Sprung am Intervallende zur Folge haben, hat jedoch einen Vorteil bei Bearbeitung mittels einer geeigneten Filterbank.

Betrachten wir dazu als einfachste Variante den Operator $\left(\downarrow\downarrow s\right)$ der Polyphasenzerlegung zu einem Faktor $s \in \mathbb{N}_{>1}$. Ist die Periodenlänge der periodischen Fortsetzung bzw. die Länge des Folgenstücks ein Vielfaches Ls von s, $L \in \mathbb{N}_{>0}$, so sind die Polyphasenfolgen dieser Fortsetzung periodisch mit Periode L. Da eine periodische Folge durch die Werte einer Periode vollständig bestimmt ist, ist die Gesamtheit der Polyphasen durch genau so viele Glieder bestimmt wie die Ausgangsfolge. Diese Eigenschaft überträgt sich auf beliebige orthogonale Analyse–Filterbänke. Bei der Fortsetzung durch Nullsetzen entstehen in derselben Situation im Allgemeinen Folgen mit einer Länge größer als L.

Um den verbleibenden Sprung an der Grenze zweier Perioden ebenfalls zu unterdrücken, kann man die periodische Fortsetzung um einen Schritt erweitern. Spiegelt man das kurze Folgenstück an einer Intervallgrenze und setzt die doppelt so lange Folge periodisch fort, so entsteht eine periodische Folge doppelter Folgenlänge ohne neu hinzukommende Sprünge. Ist die Folgenlänge ein Vielfaches von $s \in \mathbb{N}_{>1}$, und damit die Periodenlänge $2Ls$ mit einem $L \in \mathbb{N}_{>0}$, so wird diese Folge durch eine $(s, 1)$–periodische Filterbank in ein Tupel von Folgen der

Periode $2L$ transformiert. Damit die Folgen im Ergebnis der Filterbank wieder durch L Glieder vollständig bestimmt sind, muss eine Symmetrie der Filterbank vorausgesetzt werden.

3.3.1 Spiegelsymmetrien auf Vektorräumen

Seien V ein reeller endlichdimensionaler euklidischer Vektorraum und $J : V \to V$ eine *Symmetrie*. Darunter wollen wir hier verstehen, dass $J^* = J$ und $J^2 = I_V$ die Identität auf V ist. Diese Eigenschaften sind trivialerweise für $J = I_V$ erfüllt. Dann sind $\frac{1}{2}(I_V - J)$ und $\frac{1}{2}(I_V + J)$ orthogonale Projektoren in V mit Summe I_V und zueinander senkrecht stehenden Bildräumen. Diese erzeugen eine Zerlegung des Vektorraums V in einen Teilraum V_+ *gerader* Vektoren mit $v = Jv$ und einen Teilraum V_- *ungerader* Vektoren mit $v = -Jv$.

Beispielsweise kann auf einem Spaltenvektorraum $V = \mathbb{R}^d$ eine Symmetrie durch Vertauschen der Reihenfolge definiert werden, d.h.

$$v = (v_1, v_2, \ldots, v_d)^t \mapsto J(v) := (v_d, v_{d-1}, \ldots, v_1)^t \,.$$

Sei $m \in \mathbb{N}$ durch $d - 1 \leq 2m \leq d$ definiert. Dann erzeugt die Anwendung der Projektoren auf die kanonische Basis eine Basis aus $m + 1$ geraden Vektoren von V_+ und von m ungeraden Vektoren von V_-.

3.3.2 Symmetrieeigenschaften von Folgen

Auf einem beliebigen Folgenraum $\ell_{\text{fin}}(W)$ können auf einfache Weise Symmetrien durch Richtungswechsel im Index definiert werden. Für jedes $\nu \in \mathbb{Z}$ seien $\tau_\nu : \ell_{\text{fin}}(W) \to \ell_{\text{fin}}(W)$ diejenige Symmetrie, die die Folge $a = \{a_n\}_{n\in\mathbb{Z}}$ auf die am (virtuellen) Index $\nu/2$ gespiegelte Folge $\tau_\nu(a) := \{a_{\nu-n}\}_{n\in\mathbb{Z}}$ abbildet. Für jede Verzögerung ν kann die Spiegelung in eine Verschiebung um ν und eine Spiegelung am Index 0 zerlegt werden, $\tau_\nu = \tau_0 \circ \mathcal{T}^\nu$.

Definition 3.3.1 *Sei V ein $\mathbb{R}$–Vektorraum. Eine Folge $a = \{a_k\}_{k\in\mathbb{Z}} \in \ell_{\text{fin}}(V)$ heißt* symmetrisch *bzw.* antisymmetrisch *mit* Verzögerung $\nu \in \mathbb{Z}$, *wenn $\tau_\nu a = a$ bzw. $\tau_\nu a = -a$ gilt, d.h.*

$$\forall k \in \mathbb{Z} : \; a_{\nu-k} = \sigma a_k$$

mit Vorzeichen $\sigma = 1$ bzw. $\sigma = -1$ gilt.

Für die elementaren verschiebungsinvarianten Operationen gelten folgende Regeln der Indexspiegelung.

Lemma 3.3.2 *Seien V, W zwei $\mathbb{R}$–Vektorräume, $f \in \ell_{\text{fin}}\big(\text{Hom}(V, W)\big)$ eine endliche Folge linearer Abbildungen und $p \in \mathbb{N}_{>1}$. Dann gelten für jede Verzögerung $\nu \in \mathbb{Z}$*

- $f(\mathcal{T}) \circ \tau_\nu = \tau_\nu \circ f(\mathcal{T}^{-1})$,
- $\big(\uparrow\!\text{p}\big) \circ \tau_\nu = \tau_{p\nu} \circ \big(\uparrow\!\text{p}\big)$ *und*
- $\big(\downarrow\!\text{p}\big) \circ \tau_{p\nu} = \tau_\nu \circ \big(\uparrow\!\text{p}\big)$.

Beweis: Eine Verschiebung kombiniert mit einer Spiegelung am Index 0 hat den gleichen Effekt wie die Spiegelung am Index Null gefolgt von einer gleichgroßen Verschiebung in umgekehrter Richtung. Somit ist

$$f(\mathcal{T}) \circ \tau_\nu = f(\mathcal{T}) \circ \tau_0 \circ \mathcal{T}^\nu = \tau_0 f(\mathcal{T}^{-1}) \mathcal{T}^\nu = \tau_\nu \circ f(\mathcal{T}^{-1}) \,. \tag{3.2}$$

Weiter vertauscht die Spiegelung am Index Null mit dem Übertaktungsoperator, daraus folgt

$$\big(\uparrow \mathrm{p}\big) \circ \tau_\nu = \big(\uparrow \mathrm{p}\big) \circ \tau_0 \circ \mathcal{T}^\nu = \tau_0 \circ \mathcal{T}^{p\nu} \circ \big(\uparrow \mathrm{p}\big) = \tau_{p\nu} \circ \big(\uparrow \mathrm{p}\big) \,, \tag{3.3}$$

die dritte Beziehung folgt durch Übergang zur adjungierten Identität. □

3.3.3 Struktur symmetrischer Filterbänke

Wir suchen Filterbänke, die (anti–)symmetrische Folgen in (anti–)symmetrische Folgen überführen und dabei das Symmetriezentrum erhalten.

Satz 3.3.3 *Seien $\nu \in \mathbb{Z}$, $s \in \mathbb{N}_{>1}$, V ein $\mathbb{R}$–Vektorraum und $a \in \ell_{\mathrm{fin}}(\mathrm{Hom}(V,V))$ eine endliche Folge linearer Abbildungen.*

Dann gilt für den $(s,1)$–periodischen Operator $F := a(\mathcal{T}) \circ \big(\uparrow \mathrm{s}\big)$ die Identität $F \circ \tau_\nu = \tau_\nu \circ F$ genau dann, wenn a symmetrisch mit Verzögerung $(1-s)\nu$ ist.

Beweis: Nach dem vorhergehenden Lemma gilt

$$F \circ \tau_\nu = a(\mathcal{T}) \circ \tau_{s\nu} \circ \big(\uparrow \mathrm{s}\big) = \tau_{s\nu} \circ a(\mathcal{T}^{-1}) \circ \big(\uparrow \mathrm{s}\big) \,.$$

Um die gewünschte Identität zu erhalten, muss $\tau_0 \mathcal{T}^{s\nu} a(\mathcal{T}^{-1}) = \tau_0 \mathcal{T}^\nu a(\mathcal{T})$ gelten, also $a = \mathcal{T}^{s\nu-\nu} \tau_0 a = \tau_{(1-s)\nu} a$. □

In Analogie dazu muss a antisymmetrisch mit der gleichen Verzögerung sein, wenn F symmetrische in antisymmetrische Folgen überführt, ohne das Symmetriezentrum zu ändern.

3.4 Frequenzselektive Filterbänke

3.4.1 Laurent– und trigonometrische Polynome

Sind $c(\mathcal{T}) := \sum_{k\in\mathbb{Z}} c_k \mathcal{T}^k$ und $d(\mathcal{T}) := \sum_{l\in\mathbb{Z}} d_l \mathcal{T}^l$ zwei Differenzenoperatoren mit endlichen Koeffizientenfolgen $c, d \in \ell_{\mathrm{fin}}(\mathbb{C})$, so ist die Verknüpfung beider

$$c(\mathcal{T}) \circ d(\mathcal{T}) = \sum_{k=-m}^{m} c_k \mathcal{T}^k \circ \sum_{l=-n}^{n} d_l \mathcal{T}^l = \sum_{k=-(m+n)}^{m+n} \left(\sum_{l=0}^{k} c_{k-l} d_l \right) \mathcal{T}^k \,.$$

Die Koeffizienten dieser Verknüpfung sind die Glieder der Folge des Faltungsprodukts $c * d$.

Definition 3.4.1 *Sei $\mathcal{R}$ ein Ring und $\ell_{\text{fin}}(\mathbb{Z}, \mathcal{R})$ der Raum der endlichen Folgen. Die $\mathcal{R}$–Algebra $\mathcal{R}\langle Z\rangle$ der* Laurent–Polynome *besteht aus diesem Raum mit der gliedweisen Addition, gliedweisen Multiplikation mit Elementen von $\mathcal{R}$ und Multiplikation durch Faltung.*

Die Monome Z^n entsprechen den einelementigen Folgen δ^n mit Wert 1 an der Stelle n.

D.h., für jede endlich lange Folge $a = \{a_n\}_{n\in\mathbb{Z}} \in \ell_{\text{fin}}(\mathcal{R})$, aufgefasst als Laurent–Polynom, gilt $a = \sum_{k=-N}^{N} a_k Z^k$ mit einem genügend großen $N \in \mathbb{N}$. Alternativ läßt sich diese Algebra als Quotientenring (s. Definition 2.2.1) $\mathcal{R}\langle Z\rangle = \mathcal{R}[Z, Z^{-1}] := \mathcal{R}[Z, W]/\langle\, ZW - 1\,\rangle$ definieren.

Laurent–Polynome können ausgewertet werden, indem die Monome Z^n durch die entsprechenden Potenzen eines invertierbaren Elements von $\mathcal{R}$ ersetzt werden. Ist $\mathcal{R}_1$ ein weiterer Ring, so dass ein Ringhomomorphismus $\varphi : \mathcal{R} \to \mathcal{R}_1$ besteht, so kann jedes Laurent-Polynom aus $\mathcal{R}\langle Z\rangle$ in ein Laurent–Polynom in $\mathcal{R}_1\langle Z\rangle$ umgewandelt und auf invertierbaren Elementen von $\mathcal{R}_1$ ausgewertet werden.

Die Differenzenoperatoren mit komplexen Koeffizienten bilden einen Ring, in welchem die einfache Verschiebung invertierbar ist. Laurent–Polynome aus $\mathbb{C}\langle Z\rangle$ können also in $\mathcal{T}$ und jeder Potenz davon ausgewertet werden, das Bild dieser Auswertungsabbildung ist der gesamte Ring der Differenzenoperatoren.

Ein Laurent–Polynom mit komplexen Koeffizienten kann ebenso in jeder von Null verschiedenen komplexen Zahl ausgewertet werden. Nach der Theorie der Fourier–Reihen bzw. der Cauchyschen Integralformel genügt es, die Auswertung des Laurent–Polynoms auf dem Einheitskreis zu kennen, um sämtliche Koeffizienten rekonstruieren zu können. Dies läßt sich auf vektorwertige Koeffizientenfolgen verallgemeinern.

Definition 3.4.2 *Seien V ein $\mathbb{C}$–Vektorraum und $c \in \ell_{\text{fin}}(V)$ eine endliche Folge. Für jedes $\omega \in \mathbb{R}$ kann die Linearkombination*

$$\hat{c}(\omega) := \sum_{n\in\mathbb{Z}} e^{-i(2\pi\omega)n} c_n$$

gebildet werden, da nur endlich viele Glieder der Reihe von Null verschieden sind. Die so definierte Funktion $\hat{c} : \mathbb{R} \to V$ wird vektorwertiges trigonometrisches Polynom *genannt.*

Für $V = \mathbb{C}$ entspricht dieses der Auswertung des Laurent–Polynoms $c(Z)$ in $z = e^{-i2\pi\omega}$.

Aus einem trigonometrischen Polynom kann man auch wieder die Koeffizientenfolge bestimmen. Systematisch erfolgt dies, indem die Orthogonalität der trigonometrischen Monome e_n, $\omega \mapsto e_n(\omega) := e^{i2\pi\omega n}$ ausgenutzt wird. Es gilt

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi\omega n}\,\mathrm{d}\omega = \begin{cases} 1 & \text{bei } n = 0, \\ 0 & \text{bei } n \neq 0\,, \end{cases}$$

damit gewinnt man die Koeffizienten durch *Fourier–Integrale* zurück,

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{c}(\omega) e^{i2\pi\omega n}\,\mathrm{d}\omega = \sum_{k\in\mathbb{Z}} c_k \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi\omega(n-k)}\,\mathrm{d}\omega = c_n\,. \tag{3.4}$$

Man kann diese Beziehung als eine Darstellung der Folge c durch die elementaren Schwingungsfolgen $\{e_n(\omega)\}_{n\in\mathbb{Z}}$ zu Frequenzen $\omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ interpretieren. Die elementare Schwingungsfolge der Frequenz ω ist – in dieser Interpretation – in der Folge c mit dem vektorwertigen „Gewicht“ $\hat{c}(\omega)$ vertreten. Die Schwingungsfolge zur Frequenz $\omega = 0$ ist die konstante Folge mit Gliedern 1, die Schwingungsfolgen zu den Randfrequenzen $\omega = \pm\frac{1}{2}$ sind identisch und gleich der alternierenden Folge $\{(-1)^n\}_{n\in\mathbb{Z}}$. Dieses sind die extremen Möglichkeiten der Oszillationsgeschwindigkeit einer Folge.

Seien V und W zwei $\mathbb{C}$–Vektorräume und $f \in \ell_{\text{fin}}\big(\text{Hom}(V,W)\big)$ eine endliche Folge von linearen Abbildungen. Dann gibt es auch zu f ein trigonometrisches Polynom $\hat{f} : \mathbb{R} \to \text{Hom}(V,W)$.

3.4.2 Darstellung der elementaren periodischen Operatoren

Für die elementaren periodischen Operatoren gelten folgende Transformationsregeln trigonometrischer Polynome:

Satz 3.4.3 *Seien V, W zwei $\mathbb{C}$–Vektorräume, $a \in \ell_{\text{fin}}(V)$ eine endliche vektorwertige Folge, $\hat{a}$ deren trigonometrisches Polynom.*

i) Ist $f \in \ell_{\text{fin}}\big(\text{Hom}(V,W)\big)$ eine endliche Folge linearer Abbildungen und $\hat{f}$ deren trigonometrisches Polynom, so gilt für das trigonometrische Polynom der Bildfolge $b := f(\mathcal{T})(a) \in \ell_{\text{fin}}(W)$

$$\hat{b}(\omega) = \hat{f}(\omega)\big(\hat{a}(\omega)\big) \qquad \forall\omega \in \mathbb{R}\,.$$

ii) Seien $q \in \mathbb{N}_{>0}$ und $b := \big(\downarrow\mathrm{q}\big)(a)$. Dann gilt für die trigonometrischen Polynome die Beziehung

$$\hat{b}(\omega) = \frac{1}{q}\sum_{j=0}^{q-1} \hat{a}\big(\tfrac{\omega+j}{q}\big) \qquad \forall\omega \in \mathbb{R}\,.$$

iii) Seien $q \in \mathbb{N}_{>0}$ und $b := \big(\downarrow\mathrm{q}\big)(a)$. Dann gilt für die trigonometrischen Polynome die Beziehung

$$\hat{b}(\omega) = \hat{a}(s\omega) \qquad \forall\omega \in \mathbb{R}\,.$$

Beweis: *zu i)* Das Bild einer endlichen Folge $a \in \ell_{\text{fin}}(V)$ unter $f(\mathcal{T})$ ist die ebenfalls endliche Folge $b = f(\mathcal{T})(a) = \sum_{n\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} f_k(a_{n-k})\delta^n$. Dieser ist das vektorwertige trigonometrische Polynom

$$\hat{b}(\omega) = \sum_{n\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} f_k(a_{n-k})e_{-n}(\omega) = \sum_{k\in\mathbb{Z}} e_{-k}(\omega) f_k\left(\sum_{n\in\mathbb{Z}} a_{n-k}e_{k-n}(\omega)\right) = \sum_{k\in\mathbb{Z}} \hat{f}(\omega)(\hat{a}(\omega))$$

zugeordnet.

zu ii) Die Folge $b = \big(\downarrow\mathrm{q}\big)(a)$ ist eine Teilfolge von a, kann also höchstens so viele von Null verschiedene Glieder wie a besitzen. Für die Beziehungen zwischen den trigonometrischen

Polynomen muss das Weglassen der restlichen Folgenglieder in a berücksichtigt werden. Dazu können die Summen der q–ten Einheitswurzeln verwendet werden, denn es gilt

$$\sum_{j=0}^{q-1} e_n\left(\tfrac{j}{q}\right) = \sum_{j=0}^{q-1} \left(e^{i(2\pi q^{-1})n}\right)^j = \begin{cases} q & \text{wenn } n \text{ ein Vielfaches von } q \text{ ist, und} \\ 0 & \text{sonst.} \end{cases}$$

Benutzen wir dies, um die weggelassenen Folgenglieder auszublenden, so folgt

$$\begin{aligned} \hat{b}(\omega) &= \sum_{n\in\mathbb{Z}} b_n e_{-n}(\omega) = \sum_{n\in\mathbb{Z}} a_{qn} e_{-n}(\omega) \\ &= \sum_{m\in\mathbb{Z}} a_m \frac{1}{q} \sum_{j=0}^{q-1} e_{-m}\left(\tfrac{j}{q}\right) e_{-q^{-1}m}(\omega) = \frac{1}{q} \sum_{j=0}^{q-1} \sum_{m\in\mathbb{Z}} a_m e_{-m}\left(\tfrac{\omega+j}{q}\right), \end{aligned}$$

woraus sich nach Definition des vektorwertigen trigonometrischen Polynoms $\hat{a}$ die Behauptung ergibt.

zu iii) Da $b = \big(\uparrow\mathrm{q}\big)(a)$ neben den Gliedern der Folge a nur aus Nullvektoren aufgebaut ist, bleibt die Anzahl von Null verschiedener Glieder endlich. Aus demselben Grund ist $\hat{b}(\omega) = \sum_{n\in\mathbb{Z}} a_n e_{-sn} = \hat{a}(s\omega)$. □

Für die (p,q)–periodische Verknüpfung $F = \big(\downarrow\mathrm{q}\big) \circ f(\mathcal{T}) \circ \big(\uparrow\mathrm{p}\big)$ und Folgen $a \in \ell_{\mathrm{fin}}(V)$, $b := F(a) \in \ell_{\mathrm{fin}}(W)$ gilt zusammenfassend

$$\hat{b}(\omega) = \frac{1}{q} \sum_{j=0}^{q-1} \hat{f}\Big(\tfrac{\omega+j}{q}\Big) \Big(\hat{a}\Big(\tfrac{p}{q}(\omega+j)\Big)\Big) \ . \tag{3.5}$$

Sind U, V, W hermitesche $\mathbb{C}$–Vektorräume, $f \in \ell_{\mathrm{fin}}\big(\mathrm{Hom}(V,W)\big)$ und $g \in \ell_{\mathrm{fin}}\big(\mathrm{Hom}(W,U)\big)$ zwei endliche Folgen linearer Abbildungen sowie $F = f(\mathcal{T}) : \ell_{\mathrm{fin}}(V) \to \ell_{\mathrm{fin}}(W)$ und $G = g(\mathcal{T}) : \ell_{\mathrm{fin}}(W) \to \ell_{\mathrm{fin}}(U)$ die zugehörigen Differenzenoperatoren, so gilt für deren Verknüpfung $H := G \circ F$ mit Koeffizientenfolge $h \in \ell_{\mathrm{fin}}\big(\mathrm{Hom}(V,U)\big)$ für jedes $\omega \in \mathbb{R}$ die Identität der abbildungswertigen trigonometrischen Polynome $\hat{h}(\omega) = \hat{g}(\omega) \circ \hat{f}(\omega)$. Ist insbesondere $U = V$ und G linksinvers zu F, so ist $H = id_{\ell_{\mathrm{fin}}(V)}$ und für jedes $\omega \in \mathbb{R}$ gilt $\hat{h}(\omega) = id_W$, damit ist auch $\hat{g}(\omega)$ linksinvers zu $\hat{f}(\omega)$.

Gibt es die Folge $f^* = \{f^*_{-k}\}_{k\in\mathbb{Z}} \in \ell_{\mathrm{fin}}\big(\mathrm{Hom}(W,V)\big)$ der adjungierten linearen Abbildungen, so gilt $f(\mathcal{T})^* = f^*(\mathcal{T})$. Für deren abbildungswertiges trigonometrisches Polynom ergibt sich für jedes $\omega \in \mathbb{R}$

$$\widehat{f^*}(\omega) = \sum_{k\in\mathbb{Z}} f^*_{-k} e_{-k}(\omega) = \left(\sum_{k\in\mathbb{Z}} f_{-k} e_k(\omega)\right)^* = \hat{f}(\omega)^* \ .$$

Ist also $f(\mathcal{T})$ semi–unitär, so gilt dies auch für $\hat{f}(\omega)$ für jedes $\omega \in \mathbb{R}$.

3.4.3 Darstellung symmetrischer Folgen

Die definierenden Beziehungen einer symmetrischen oder antisymmetrischen Folge stellen Abhängigkeiten zwischen ihren Koeffizienten dar. Für eine im Index Null symmetrische Folge $a \in \ell_{\text{fin}}(V)$ gilt $a_{-n} = a_n$. Für das trigonometrische Polynom dieser Folge ergibt sich somit

$$\hat{a}(\omega) = a_0 + \sum_{k=1}^{\infty} a_k(e_{-k}(\omega) + e_k(\omega)) = a_0 + 2\sum_{k=1}^{\infty} a_k \cos(2\pi\, k\omega)\ .$$

Lemma 3.4.4 *Sei V ein $\mathbb{C}$–Vektorraum und $a \in \ell_{\text{fin}}(V)$ eine endliche, zum Index Null symmetrische Folge. Dann gibt es eine endliche Folge $b = \{b_k\}_{k\in\mathbb{N}_0} \in \ell_{\text{fin}}(\mathbb{N}, V)$, mit welcher sich das trigonometrische Polynom von a als polynomialer Ausdruck in $\cos(2\pi\omega)$ darstellen läßt,*

$$\hat{a}(\omega) = b(\cos(2\pi\omega)) := \sum_{k=0}^{\infty} b_k \cos(2\pi\omega)^k\ .$$

Beweis: Dies folgt aus den trigonometrischen Additionstheoremen und findet eine Formulierung in der Existenz der Tschebyscheff–Polynome $T_k \in \mathbb{Q}[X]$. Mit diesen gilt für jedes $n \in \mathbb{Z}$ die Identität $\cos n\alpha = T_n(\cos(a))$. Die Tschebyscheff–Polynome sind rekursiv definiert durch $T_0 = 1$, $T_1 = X$ und $T_{n+1} = 2T_nX - T_{n-1}$ für alle $n \in \mathbb{N}_{>0}$. Letztere Rekursionsvorschrift folgt aus der Summation der beiden Identitäten

$$\cos((n \pm 1)\alpha) = \cos n\alpha \cos\alpha \mp \sin n\alpha \sin\alpha\ .$$

Es gilt somit $\hat{a}(\omega) = a_0 + 2\sum_{k=1}^{\infty} a_k T_k(\cos(2\pi\omega))$, durch Zusammenfassen gleicher Potenzen ergibt sich die Folge b. □

Entwickelt man das Polynom b im Punkt 1 und beachtet, dass für endliche Folgen vom trigonometrischen Polynom zum Laurent–Polynom übergegangen werden kann, so ergibt sich als Folgerung:

Korollar 3.4.5 *Für jede endliche, zum Index Null symmetrische symmetrische Folge $a \in \ell_{\text{fin}}(\mathbb{C})$ gibt es ein Polynom $b \in \mathbb{C}[X]$, so dass für das Laurent–Polynom zu a gilt*

$$a(Z) = b\big(1 - \tfrac{1}{2}(Z + Z^{-1})\big)\ .$$

Satz 3.4.6 *Seien $a \in \ell_{\text{fin}}(\mathbb{C})$ eine symmetrische Folge mit Verzögerung $d \in \mathbb{Z}$, $a = \tau_d a$ und $X := 1 - \frac{1}{2}(Z + Z^{-1}) \in \mathbb{Q}\langle Z\rangle$ ein symmetrisches Laurent–Polynom. Dann gibt es ein Polynom $b \in \mathbb{C}[X]$, so dass das Laurent–Polynom $a(Z) \in \mathbb{C}\langle Z\rangle$*

- *bei gerader Verzögerung $d = 2m$ eine Darstellung*

$$a(Z) = Z^m b(X),$$

- *bei ungerader Verzögerung $d = 2m + 1$ eine Darstellung*

$$a(Z) = Z^m(1 + Z)b(X)$$

besitzt.

Beweis: Bei gerader Verzögerung gilt $a(Z) = Z^{2m}a(Z^{-1})$, also hat $Z^{-m}a(Z)$ Verzögerung 0 und damit eine Parametrisierung $Z^{-m}a(Z) = b(X)$.

Bei ungerader Verzögerung ist $z = -1$ eine Nullstelle von $a(Z)$, denn es gilt $a(-1) = -a(-1)$. Nach Abspalten des Faktors $(Z+1)$ der Verzögerung 1 erhalten wir ein symmetrisches Polynom gerader Verzögerung $2m$. □

3.4.4 Tiefpassfilter und Haar–Polynom

Für die weitere Entwicklung der Interpretation des vektorwertigen trigonometrischen Polynoms als Gewichtsfunktion elementarer Schwingungsfolgen sei V mit einer Norm $\|\cdot\|$ ausgestattet; für $V = \mathbb{C}$ kann der Absolutbetrag als Norm gewählt werden. Man kann nun Folgen danach bewerten, ob sie eher dem „langsamen" Typ der konstanten Folge oder dem „schnellen" Typ der alternierenden Folge entsprechen. D.h. danach, ob die Funktion $\|\hat{c}\|$, die ja periodisch mit Periode 1 ist, ihre Maxima eher nahe Null oder eher in der Nähe von $\pm\frac{1}{2}$ besitzt. Je nach Anwendungsgebiet muss diese vage Begriffsbildung in genauen Definitionen konkretisiert werden. Uns genügt im Folgenden die Betrachtung des lokalen Verhaltens von $\hat{c}$ in 0 bzw. in den extremen Frequenzen $\pm\frac{1}{2}$.

Seien V und W zwei $\mathbb{C}$–Vektorräume, $p, q \in \mathbb{N}_{>0}$ und $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ ein (p,q)-periodischer Operator, der durch eine endliche Folge von Koeffizientenabbildungen $f \in \ell_{\text{fin}}\big(\text{Hom}(V,W)\big)$ als $F = \big(\downarrow\text{q}\big) \circ f(\mathcal{T}) \circ \big(\uparrow\text{p}\big)$ gegeben ist. Man nennt F ein *Tiefpassfilter*, wenn die Bildfolgen von F eher dem langsamen Typ angehören. Insbesondere soll eine langsame Folge a wieder eine langsame Folge $b = F(a)$ als Bild haben. Nach Gleichung (3.5) gilt

$$\hat{b}\left(\omega + \tfrac{kq}{p}\right) = \tfrac{1}{q}\hat{f}\left(\tfrac{k}{p} + \tfrac{\omega}{q}\right)\left(\hat{a}\left(\tfrac{p\omega}{q}\right)\right) + \tfrac{1}{q}\sum_{j=1}^{q-1}\hat{f}\left(\tfrac{k}{p} + \tfrac{\omega+j}{q}\right)\left(\hat{a}\left(\tfrac{p(\omega+j)}{q}\right)\right) .$$

Die langsame Komponente von a soll einen Einfluss auf die langsame Komponente von b haben. Betrachtet man den Fall $k = 0$ und $\omega \approx 0$, so ergibt sich als Bedingung, dass $\hat{f}(0)$ von Null verschieden sei. Hat die Norm $\|\hat{a}(\omega)\|$ ein Maximum in $\omega = 0$ und ist außerhalb des Intervalls $\left[-\frac{1}{2q}, \frac{1}{2q}\right]$ klein relativ zum Maximum, so ergibt sich für $\omega \approx 0$ und $k = 1, \ldots, p-1$ der wesentliche Anteil an $\hat{b}\left(\omega + \frac{kq}{p}\right)$ aus dem ersten Summanden. Um diesen Anteil klein zu halten, kann $\hat{f}(\frac{k}{p}) = 0$ für $k = 1, \ldots, p-1$ gefordert werden.

Seien die Folgen komplexwertig, es gelte also $V = W = \mathbb{C}$ und damit auch $\text{Hom}(V,W) = \mathbb{C}$. Dann kann die Nullstellenbedingung an das trigonometrische Polynom $\hat{f} : \mathbb{R} \to \mathbb{C}$ genauer gefasst werden: Hat $\hat{f}$ Nullstellen in den Punkten $\frac{k}{p}$, $k = 1, \ldots, p-1$, so hat das Laurent–Polynom $f(Z) = \sum_{k\in\mathbb{Z}} f_k Z^k$ Nullstellen in den nichttrivialen p–ten Einheitswurzeln $e^{i(2\pi p^{-1})k}$, $k = 1, \ldots, p-1$. Das Minimalpolynom dieser Nullstellen ist

$$H_p(Z) := \tfrac{1}{p}\left(1 + Z + \cdots + Z^{p-1}\right) = \frac{1}{p}\,\frac{Z^p - 1}{Z - 1} . \tag{3.6}$$

Der Vorfaktor wurde so gewählt, dass $H_p(1) = 1$ gilt.

Definition 3.4.7 *Für jedes $p \in \mathbb{N}_{>1}$ wird $H_p(Z)$* Haar–Polynom *der Ordnung p genannt.*

Die Vielfachheit, mit welcher das Haar–Polynom $H_p(Z)$ in einem beliebigen Laurent–Polynom $f(Z)$, $f \in \ell_{\text{fin}}(\mathbb{C})$, vorkommt, heißt polynomiale Approximationsordnung *von f, bzw. des mit f gebildeten (p,q)–periodischen Operators $(\downarrow\mathrm{q})\, f(\mathcal{T})\, (\uparrow\mathrm{p})$ bei beliebigem $q \in \mathbb{N}_{>1}$.*

Der Begriff der polynomialen Approximationsordnung hat seinen Ursprung in der Betrachtung *polynomialer Folgen*, d.h. von Folgen $a \in \ell(\mathbb{C})$, zu denen es ein Polynom $p \in \mathbb{C}[X]$ gibt mit $a = \{p(n)\}_{n\in\mathbb{Z}}$. Sei der Raum $P_m(\mathbb{C})$ der polynomialen Folgen zum Grad m definiert als

$$P_m(\mathbb{C}) := \{\{p(n)\}_{n\in\mathbb{Z}} : p \in \mathbb{C}[X] \;\&\; \deg p \leq m\}\,.$$

Dieser Raum ist eindeutig charakterisiert als Raum aller Folgen, die durch den Differenzenoperator $(1-\mathcal{T})^{m+1}$ auf die Nullfolge abgebildet werden.

Lemma 3.4.8 *Seien $p, q \in \mathbb{N}_{>1}$ und $f \in \ell_{\text{fin}}(\mathbb{C})$ eine Folge, für die der Operator $F := (\downarrow\mathrm{q})\, f(\mathcal{T})\, (\uparrow\mathrm{p})$ eine Approximationsordnung $A \in \mathbb{N}_{>0}$ hat. Dann bildet F jeden Raum $P_m(\mathbb{C})$ polynomialer Folgen mit $m < A$ auf sich selbst ab.*

Beweis: Nach Voraussetzung gibt es eine Folge $g \in \ell_{\text{fin}}(\mathbb{C})$ mit $f(Z) = H_p(Z)^A g(Z)$. Sei $a \in P_m(\mathbb{C})$ mit $m < A$ eine polynomiale Folge; es gilt also $(1-\mathcal{T})^{m+1} c = 0$. Zu prüfen ist das Verschwinden von $(1-\mathcal{T})^{m+1} F(a)$. Wegen $(1-Z)H_p(Z) = \frac{1}{p}(1-Z^p)$ und $m+1 \leq A$ gilt für diesen Ausdruck

$$\begin{aligned}
(1-\mathcal{T})^{m+1} F(a) =& (1-\mathcal{T})^{m+1}((\downarrow\mathrm{q})\, f(\mathcal{T})\, (\uparrow\mathrm{p}))(c) \\
=& (\downarrow\mathrm{q}) \left(q^{m+1} H_q(\mathcal{T})^{m+1} (1-\mathcal{T})^{m+1} f(\mathcal{T})\, (\uparrow\mathrm{p})(c)\right) \\
=& \left(\tfrac{q}{p}\right)^{m+1} (\downarrow\mathrm{q}) \left(H_q(\mathcal{T})^{m+1}\, g_m(\mathcal{T})\, (1-\mathcal{T}^p)^{m+1}\, (\uparrow\mathrm{p})(c)\right) \\
=& \left(\tfrac{q}{p}\right)^{m+1} (\downarrow\mathrm{q}) \left(H_q(\mathcal{T})^{m+1}\, g_m(\mathcal{T})\right)\, (\uparrow\mathrm{s}) \left((1-\mathcal{T})^{m+1} c\right)\,.
\end{aligned}$$

mit $g_m(Z) = H_p(Z)^{A-m-1}\, g(\mathcal{T})$. Da nach Voraussetzung der letzte Faktor schon die Nullfolge ist, gilt das für den gesamten Ausdruck. □

3.5 Wavelet–Filterbänke

Ist eine sich nur langsam ändernde Folge als Signal vorgegeben, so kann man versuchen, diese mittels einer durch ein Tiefpassfilter erzeugten Folge zu approximieren. Ist das Tiefpassfilter $(s,1)$–periodisch, so ist das Argument eine Folge geringerer Datenrate. Hat die Approximation eine ausreichende Qualität, so kann man von einer Kompression der gegebenen Folge sprechen. Reicht die Qualität der Approximation nicht aus, so muss eine auf gleiche Art konstruierte Korrektur hinzugefügt werden. Jedoch muss das verwendete Filter kein Tiefpassfilter sein. Dieses Zusammenspiel eines Tiefpassfilters mit einem Korrekturfilter ist die Grundstruktur einer *Wavelet–Synthese–Filterbank*. Um die Korrektheit der Approximation zu sichern, wird deren Rechtsinvertierbarkeit verlangt.

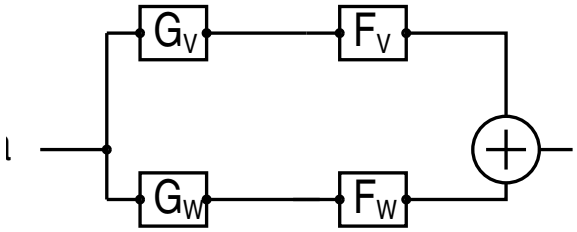
3.5.1 Wavelet–Filterbänke und Kaskaden–Schema

Definition 3.5.1 *Seien V, W endlichdimensionale $\mathbb{C}$–Vektorräume und $V \oplus W$ deren direkte Summe. Sei weiter $s \in \mathbb{N}_{>1}$. Ein $(1,s)$–periodischer linearer Operator $F : \ell_{\text{fin}}(V \oplus W) \to \ell_{\text{fin}}(V)$, welcher einen rechtsinversen Operator $G : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(V \oplus W)$ besitzt, heißt* Wavelet–Synthese–Filterbank *oder kürzer* Wavelet–Filterbank. *G heißt dann* Wavelet-Analyse-Filterbank.

Ist G rechtsinvers zu F, so kann dies in den partiellen Abbildungen ausgedrückt werden, für jede Folge $a \in \ell_{\text{fin}}(V)$ gilt

$$\begin{aligned}(F_V \circ G_V + F_W \circ G_W)(a) =& F_V(G_V(a)) + F_W(G_W(a)) \\ =& F(G_V(a) \oplus G_W(a)) = F(G(a)) = a\,. \end{aligned} \tag{3.7}$$

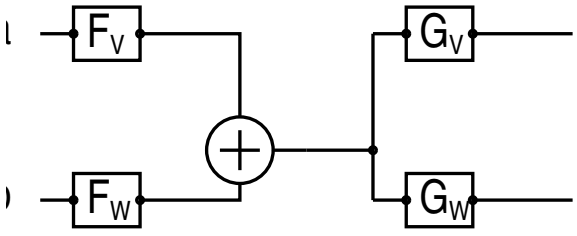
Als Schaltplan dargestellt findet sich das Signal am Eingang links unverändert am Ausgang rechts wieder.



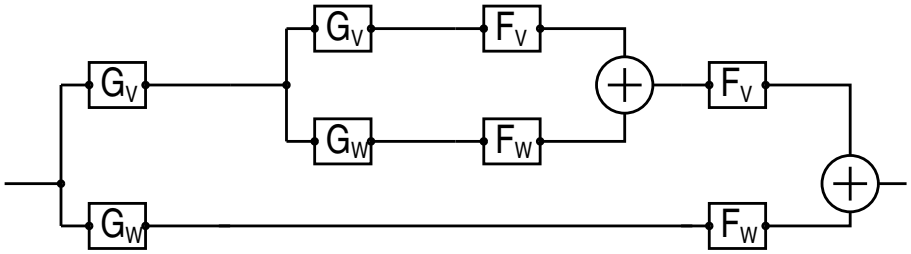
Ist G sogar invers zu F, d.h. gilt auch $G \circ F = id_{\ell_{\text{fin}}(V \oplus W)}$, so ergeben sich zusätzlich die Beziehungen

$$G_V \circ F_V = id_{\ell_{\text{fin}}(V)}, \quad G_W \circ F_W = id_{\ell_{\text{fin}}(W)}, \quad G_W \circ F_V = 0 \quad \text{und} \quad G_V \circ F_W = 0.$$

Auch diese Beziehungen kann man als Schaltplan darstellen, die Signale an den Eingängen links finden sich unverändert an den Ausgängen rechts wieder.



Versteht man die Aufspaltung mittels G und die Rekonstruktion mittels F als Aufspaltung eines Übertragungskanals des „Querschnitts“ V in zwei Unterkanäle der „Querschnitte“ V und W mit um den Faktor s reduzierter Übertragungsrate, so kann der Unterkanal mit „Querschnitt“ V wiederum in Unterkanäle aufgespalten werden.



Diese Aufteilung in Unterkanäle kann beliebig oft wiederholt werden. Es entsteht auf der linken Seite der Analyse–Operatoren eine Folge linearer Operatoren

$$\begin{aligned} G^{(1)} :=& (G_V, G_W) \\ G^{(2)} :=& (G^{(1)} \circ G_V, G_W) = (G_V \circ G_V, G_W \circ G_V, G_W) \\ G^{(k+1)} :=& (G^{(k)} \circ G_V, G_W) = (G_V^{k+1}, G_W \circ G_V^k, \ldots, G_W \circ G_V, G_W) \end{aligned} \tag{3.8}$$

und auf der rechten Seite der Synthese–Operatoren eine Folge linearer Operatoren

$$\begin{aligned} F^{(1)} :=& (F_V, F_W) \\ F^{(2)} :=& (F_V \circ F^{(1)}, F_F) = (F_V \circ F_V, F_V \circ F_W, F_W) \\ F^{(k+1)} :=& (F_V \circ F^{(k)}, F_W) = (F_V^{k+1}, F_V^k \circ F_W, \ldots, F_V \circ F_W, F_W)\,. \end{aligned} \tag{3.9}$$

Verknüpft man zu einem festen $k \in \mathbb{N}$ die Operatoren von $F^{(k)}$ komponentenweise mit denen von $G^{(k)}$ und summiert die entstehenden Abbildungen, so ergibt sich nach Konstruktion die identische Abbildung auf $\ell_{\text{fin}}(V)$.

Diese Art der Aufteilung eines Signals auf Unterkanäle bildet die Grundstruktur der diskreten Wavelet–Transformation. Die allgemeine Zielstellung bei der Auswahl des Operators F ist, dass die partielle Abbildung F_V als Tiefpassfilter und die Abbildung F_W als Hochpassfilter wirkt. Sind $c \in \ell_{\text{fin}}(V)$ und $d \in \ell_{\text{fin}}(W)$, so soll also in der Summe $F(c \oplus d) = F_V(c) + F_W(d)$ die Folge $F_V(c)$ den sich langsam ändernden Anteil und $F_W(d)$ den sich schnell ändernden Anteil beisteuern.

3.5.2 Biorthogonale Wavelet–Filterbänke

Der zu einem (p,q)–periodischen Operator adjungierte Operator wird auch *duale Filterbank* genannt. Seien F eine Wavelet–Synthese–Filterbank und G eine rechtsinverse Abbildung dazu. Dann ist die duale Filterbank G^* eine Synthese–Filterbank und F^* die rechtsinverse Analysis–Filterbank. Insofern ist es möglich, eine Analysis–Filterbank als duale Filterbank $\tilde{F}^*$ einer Synthese–Filterbank $\tilde{F}$ anzugeben. Im Rahmen der *Multiskalenanalyse* (s. Abschnitt 5.3) ist dies die natürliche Sichtweise.

Definition 3.5.2 *Seien V, W zwei $\mathbb{C}$–Vektorräume und $s \in \mathbb{N}_{>1}$. Ein Paar $(\tilde{F}, F)$ von $(1,s)$–periodischen Synthese–Filterbänken wird* biorthogonal *genannt, wenn $s^{-1}\tilde{F}^*$ invers zu F ist.*

Eine $(1,s)$–periodische Wavelet–Filterbank $F : \ell_{\text{fin}}(V \oplus W) \to \ell_{\text{fin}}(V)$ wird orthogonal *genannt, wenn das Paar (F,F) biorthogonal ist.*

Insbesondere ist für eine orthogonale Wavelet–Filterbank F der skalierte Operator $\sqrt{s^{-1}}F$ unitär. Aus der Dimensionsbetrachtung unitärer Operatoren folgt, dass $\dim V + \dim W = s \dim V$ gelten muss, d.h. $\dim W = (s-1)\dim V$. Diese Beziehung gilt auch für biorthogonale Paare von Wavelet–Filterbänken.

Der zusätzliche Faktor wurde gewählt, um die Herauftaktung durch die Synthese–Filterbank zu berücksichtigen. Sei die „Energie“ einer Folge $a \in \ell_{\text{fin}}(V)$ mit Zeittakt T als $T\|a\|^2 = \sum_{n\in\mathbb{Z}} \|a_n\|^2\, T$ definiert. Dann erhält eine orthogonale Wavelet–Filterbank diese Energie, denn die transformierte Folge $F(a \oplus 0)$ hat Zeittakt $s^{-1}T$ und damit die „Energie“

$$\|F(a\oplus b)\|^2\, s^{-1}T = \|a\oplus b\|^2 T = \|a|^2 T + \|b\|^2 T\,.$$

Jede $(s,1)$–periodische lineare Abbildung kann mittels einer endlichen Folge linearer Abbildungen angegeben werden. Im Fall einer Wavelet–Filterbank $F : \ell_{\text{fin}}(V\oplus W) \to \ell_{\text{fin}}(V)$ kann jeder der beiden Teilabbildungen F_V und F_W eine Folge $a = \{a_k\} \in \ell_{\text{fin}}\big(\text{Hom}(V,V)\big)$ bzw. $b = \{b_k\} \in \ell_{\text{fin}}\big(\text{Hom}(W,V)\big)$ zugeordnet werden. Für jedes Paar von Folgen $c \in \ell_{\text{fin}}(V)$ und $d \in \ell_{\text{fin}}(W)$ gilt dann

$$F(c\oplus d) = F_V(c) + F_W(d) = a(\mathcal{T})\,\big(\uparrow\text{s}\big)\,c + b(\mathcal{T})\,\big(\uparrow\text{s}\big)\,d\,.$$

Sei $\tilde F$ eine zweite Wavelet–Filterbank, parametrisiert durch zwei Folgen $\tilde a \in \ell_{\text{fin}}\big(\text{Hom}(V,V)\big)$ und $\tilde b \in \ell_{\text{fin}}\big(\text{Hom}(W,V)\big)$, so dass das Paar $(\tilde F, F)$ biorthogonal ist. Dann gilt für jedes Paar von Folgen $c \in \ell_{\text{fin}}(V)$ und $d \in \ell_{\text{fin}}(W)$

$$s(c\oplus d) = \big(\downarrow\text{s}\big)\,\big(\tilde a(\mathcal{T})^*\oplus\tilde b(\mathcal{T})^*\big)\,\big(a(\mathcal{T})\,\big(\uparrow\text{s}\big)\,c + b(\mathcal{T})\,\big(\uparrow\text{s}\big)\,d\big)\;.$$

Insbesondere muss, wenn man die Tiefpassanteile vergleicht,

$$\big(\downarrow\text{s}\big)\,\tilde a(\mathcal{T})^* a(\mathcal{T})\,\big(\uparrow\text{s}\big) = s\,id_{\ell_{\text{fin}}(V)} \tag{3.10}$$

gelten. Bezeichnen wir die Polyphasenteilfolgen der Folgen $a, \tilde a \in \ell_{\text{fin}}\big(\text{Hom}(V,V)\big)$ mit $\big(\downarrow\downarrow\text{s}\big)\,a = (a_{(0)},\dots,a_{(s-1)})^t$ und $\big(\downarrow\downarrow\text{s}\big)\,\tilde a = (\tilde a_{(0)},\dots,\tilde a_{(s-1)})^t$, so kann die Identität (3.10) weiter umgeschrieben werden zu

$$s\,id_{\ell_{\text{fin}}(V)} = \sum_{j=0}^{s-1} \tilde a_{(j)}(\mathcal{T})^*\,\big(\downarrow\text{s}\big)\,\mathcal{T}^{-j}\sum_{k=0} s-1\mathcal{T}^k\,\big(\uparrow\text{s}\big)\,a_{(k)}(\mathcal{T}) \tag{3.11}$$

$$= \sum_{j=0}^{s-1} \tilde a_{(j)}(\mathcal{T})^* a_{(j)}(\mathcal{T})\,, \tag{3.12}$$

denn $\big(\downarrow\text{s}\big)\,\mathcal{T}^{k-j}\,\big(\uparrow\text{s}\big)$ ist hier nur für $k=j$ von Null verschieden.

Im speziellen Fall $V=\mathbb{C}$, $W=\mathbb{C}^{s-1}$ ergeben sich daraus Beziehungen für die komplexwertigen Folgen $a, \tilde a \in \ell_{\text{fin}}(\mathbb{C})$

$$\sum_{n\in\mathbb{Z}} \bar{\tilde a}_n a_{n+sm} = \delta_{0,m}\,.$$

Sind die Glieder der Folgen a und $\tilde a$ reell, so ist dies ein System quadratischer Gleichungen.

3.5.3 Biorthogonale Paare von Skalierungsfolgen

Sei ein Skalenfaktor $s \in \mathbb{N}_{>0}$ fixiert. Wir betrachten im Folgenden nur den Fall $V = \mathbb{C}$ und damit $W = \mathbb{C}^{s-1}$. Der Tiefpassanteil einer Wavelet–Synthese–Filterbank $F : \ell_{\text{fin}}(\mathbb{C} \oplus \mathbb{C}^{s-1}) \to \ell_{\text{fin}}(\mathbb{C})$ hat die Gestalt $F_V = a(\mathcal{T})\,(\uparrow\!\mathrm{s})$ mit einer endlichen Folge $a \in \ell_{\text{fin}}(\mathbb{C})$. Diese Folge a wird als *Skalierungsfolge* der Wavelet–Synthese–Filterbank bezeichnet.

Dass der $(1,s)$–periodische Operator $F_V = a(\mathcal{T})\,(\uparrow\!\mathrm{s})$ einen Tiefpassfilter darstellt, ist nach den Überlegungen aus Abschnitt 3.4.4 in der Weise zu konkretisieren, dass dieser Operator eine polynomiale Approximationsordnung aufweist, d.h. dass das Laurent–Polynom $a(Z)$ das Haar–Polynom $H_s(Z)$ mindestens einmal als Faktor enthält. Es gibt also eine Zahl $A \in \mathbb{N}_{>0}$ und eine endliche Folge $p \in \ell_{\text{fin}}(\mathbb{C})$ mit $a(Z) = sH_s(Z)^A p(Z)$.

Seien durch $A, \tilde{A} \in \mathbb{N}_{>0}$ und $p, \tilde{p} \in \ell_{\text{fin}}(\mathbb{C})$ zwei Skalierungsfolgen $a, \tilde{a}$ mit $a(Z) = sH_s(Z)^A p(Z)$ und $\tilde{a}(Z) = sH_s(Z)^{\tilde{A}} p(Z)$ gegeben. Gehören diese zu einem biorthogonalen Paar von Wavelet–Filterbänken, folgt aus der Biorthogonalitätsbedingung für die trigonometrischen Polynome $\hat{a}$ und $\hat{\tilde{a}}$ nach den Rechenregeln von Satz 3.4.3 die Beziehung

$$s = \frac{1}{s} \sum_{j=0}^{s-1} \overline{\hat{\tilde{a}}(\tfrac{j}{s} + \omega)} \hat{a}(\tfrac{j}{s} + \omega)\,,$$

und damit auch

$$1 = \sum_{j=0}^{s-1} \overline{\hat{H}_s(\tfrac{j}{s} + \omega)^{\tilde{A}} \hat{\tilde{p}}(\tfrac{j}{s} + \omega)} \hat{H}_s(\tfrac{j}{s} + \omega)^A \hat{p}(\tfrac{j}{s} + \omega)\,.$$

Nach denselben Rechenregeln kann diese Beziehung nun als eine Bedingung an die Folge $P := \tilde{p}(\mathcal{T})^* p = \{\sum_{k \in \mathbb{Z}} \bar{\tilde{p}}_k p_{n+k}\}_{n \in \mathbb{Z}}$ formuliert werden, für diese muss

$$s\delta^0 = (\downarrow\!\mathrm{s}) \left(H_s(\mathcal{T}^{-1})^{\tilde{A}} H_s(\mathcal{T})^A P \right)$$

gelten. Diese Beziehung zwischen Folgen kann als inhomogenes lineares Gleichungssystem in den Gliedern von P aufgefasst werden. Sei $C_{s,A,\tilde{A}} : \ell_{\text{fin}}(\mathbb{C}) \to \ell_{\text{fin}}(\mathbb{C})$ die $\mathbb{C}$–lineare Abbildung

$$Q \mapsto C_{s,A,\tilde{A}}(Q) := (\downarrow\!\mathrm{s})(H_s(\mathcal{T})^A H_s(\mathcal{T}^{-1})^{\tilde{A}} Q)\,,$$

des Raums endlicher Folgen in sich. Dann ist das gesuchte Laurent–Polynom P eine Lösung des linearen Gleichungssystems $C_{s,A,\tilde{A}}(P) = s\delta^0$. Zu diesem Gleichungssystem suchen wir nun eine möglichst einfache Lösung des inhomogenen Systems und eine Charakterisierung des homogenen Lösungsraums (s. [Hel95], [BW99]).

Satz 3.5.3 *Seien ein Skalenfaktor $s \in \mathbb{N}_{>1}$ und eine Ordnung $A \in \mathbb{N}_{>0}$ gegeben. Dann gibt es für jede endliche Folge $q \in \ell_{\text{fin}}(\mathbb{C})$ mit Träger im Segment $\{0, \ldots, A-1\} \subset \mathbb{Z}$ genau eine endliche Folge $p \in \ell_{\text{fin}}(\mathbb{C})$ mit Träger im gleichen Segment, welche das lineare Gleichungssystem*

$$C_{s,A}(p) := (\downarrow\!\mathrm{s}) \left(H_s(\mathcal{T})^A p \right) = q$$

erfüllt. Jede andere Lösung dieser Gleichung hat die Form $p + (1 - \mathcal{T})^A R$, R eine weitere endliche Folge mit $(\downarrow\!\mathrm{s})\, R = 0$.

Die Folge p ist eindeutig durch die Identität $H_s(Z)^A p(Z) = sq(Z^s) \mod (1 - Z)^A$ der Laurent–Polynome bestimmt.

Beweis: Wir können jedes Laurent–Polynom $p \in \mathbb{C}\langle Z\rangle$ mittels Polynomdivision mit Rest in ein Polynom $p_0 \in \mathbb{C}[Z]$ vom Grad kleiner A und ein Vielfaches von $(Z-1)^A$ aufspalten, $p(Z) = p_0(Z) + (1-Z)^A R(Z)$. Für die Koeffizientenfolgen gilt also $p = p_0 + (1-\mathcal{T})^A R$, in das Gleichungssystem eingesetzt ergibt sich

$$C_{s,A}(p) = \big(\downarrow\mathrm{s}\big)\left(H_s(\mathcal{T})^A p\right) = \big(\downarrow\mathrm{s}\big)\left(H_s(\mathcal{T})^A p_0\right) + s^{-A}(1-\mathcal{T})^A\,\big(\downarrow\mathrm{s}\big)\,R\,.$$

Die Folge $H_s(\mathcal{T})^A p_0$ im ersten Summanden hat ihren Träger im Segment $\{0,\dots,(s-1)A + A - 1\}$. Die obere Grenze ist also $sA-1$, nach der Untertaktung verbleibt daher eine Folge mit Träger im Segment $\{0,\dots,A-1\}$. Hatte als p schon einen auf dieses Segment beschränkten Träger, d.h. ist $R = 0$, so bildet $C_{s,A}$ auf eine Folge mit Träger in diesem Segment ab. Ist R von Null verschieden, und auch $\big(\downarrow\mathrm{s}\big)\,R$ von Null verschieden, so hat $(1-\mathcal{T})^A\,\big(\downarrow\mathrm{s}\big)\,R$ ein von Null verschiedenes Glied mit Index oberhalb $A-1$. Die Folge $C_{s,A}(p)$ kann also nur einen auf $\{0,\dots,A-1\}$ beschränkten Träger haben, wenn $\big(\downarrow\mathrm{s}\big)\,R = 0$ gilt.

Sei eine Folge $q \in \ell_{\mathrm{fin}}(\mathbb{C})$ mit $\operatorname{supp} q \subset \{0,\dots,A-1\}$ gegeben. Gibt es irgendein $p \in \ell_{\mathrm{fin}}(\mathbb{C})$ mit $C_{s,A}(p) = q$, und zerlegt man dieses per Division mit Rest, so dass $p = p_0 + (1-\mathcal{T})^A R$ gilt mit $\operatorname{supp} p_0 \subset \{0,\dots,A-1\}$, so muss $\big(\downarrow\mathrm{s}\big)\,R = 0$ und damit $C_{s,A}(p_0) = q$ gelten. Für jede beliebige Folge $R \in \ell_{\mathrm{fin}}(\mathbb{C})$ mit $\big(\downarrow\mathrm{s}\big)\,R = 0$ gilt dann $C_{s,A}(p_0 + (1-\mathcal{T})^A R) = q$. Es verbleibt zu zeigen, dass es ein $p \in \ell_{\mathrm{fin}}(\mathbb{C})$ mit $\operatorname{supp} p \subset \{0,\dots,A-1\}$ und $C_{s,A}(p) = q$ gibt. Da man in diesem Zusammenhang $C_{s,A}$ als lineare Abbildung $C_{s,A} : \mathbb{C}^A \to \mathbb{C}^A$, d.h. als $A \times A$–Matrix auffassen kann, ist die Surjektivität äquivalent zur Injektivität. Insbesondere ist dann der Kern der Folgenabbildung $\ker C_{s,A} = \{R \in \ell_{\mathrm{fin}}(\mathbb{C}) :\ \big(\downarrow\mathrm{s}\big)\,R(Z) = 0\}$.

Wenn $p \in \ell_{\mathrm{fin}}(\mathbb{C})$ der Gleichung $q = C_{s,A}(p) = \big(\downarrow\mathrm{s}\big)\,\big(H_s(\mathcal{T})^A p\big)$ genügt, dann gilt nach Satz 3.4.3 für die trigonometrischen Polynome

$$\hat{q}(s\omega) = \frac{1}{s}\sum_{j=0}^{s-1} \hat{H}_s\big(\omega + \tfrac{j}{s}\big)^A\,\hat{p}\big(\omega + \tfrac{j}{s}\big)\,.$$

Auf die Laurent–Polynome übertragen bedeutet dies

$$q(Z^s) = \frac{1}{s}\sum_{j=0}^{s-1} H_s(w^j Z)^A\,p(w^j Z)\,. \tag{3.13}$$

Dabei ist $w = e^{i2\pi s^{-1}}$ eine primitive Einheitswurzel der Ordnung s. Das Polynom $H_s(Z)$ des Grades $(s-1)$ hat Nullstellen an den Einheitswurzeln $w, w^2, \dots, w^{s-1}$. Damit haben auch die Polynome $H_s(w^j Z)^A$ für $j = 0,\dots,s-1$ nur Nullstellen an den Einheitswurzeln, aber keine davon gemeinsam. Nach dem univariaten Nullstellensatz gibt es also Polynome $g_0,\dots,g_{s-1} \in \mathbb{C}[Z]$, mit welchen

$$q(Z^s) = \frac{1}{s}\sum_{k=0}^{s-1} H_s(w^k Z) g_k(w^k Z)$$

gilt. Aus diesen bilden wir das gemittelte Polynom $p(Z) := \frac{1}{s}\sum_{l=0}^{s-1} g_l(Z)$. Mit diesem können

wir jedes der g_k ersetzen, denn

$$\begin{aligned}\frac{1}{s}\sum_{k=0}^{s-1} H_s(w^kZ)^A p(w^kZ) =& \frac{1}{s^2}\sum_{k,l=0}^{s-1} H_s(w^kZ)^A g_l(w^kZ) = \frac{1}{s^2}\sum_{k,l=0}^{s-1} H_s(w^{k+l}Z)^A g_l(w^{k+l}Z)\\ =& \frac{1}{s}\sum_{k=0}^{s-1}\frac{1}{s}\sum_{l=0}^{s-1} H_s(w^l(w^kZ))^A g_l(w^l(w^kZ)) = \frac{1}{s}\sum_{k=0}^{s-1} q((w^kZ)^s)\\ =& q(Z^s)\end{aligned}$$

Somit kennen wir eine Lösung p des inhomogenen linearen Gleichungssystems $C_{s,A}(p) = q$. Mittels Polynomdivision mit Rest erhalten wir daraus, wie am Anfang des Beweises ausgeführt, eine Folge $p_0 \in \ell_{\text{fin}}(\mathbb{C})$ mit Träger in $\{0,\ldots,A-1\}$ und $C_{s,A}(p_0) = q$. Somit ist die eingeschränkte lineare Abbildung $C_{s,A} : \mathbb{C}^A \to \mathbb{C}^A$ bijektiv.

Weiterhin gilt für jede Lösung $p \in \ell_{\text{fin}}(\mathbb{C})$ durch Umstellen von Gleichung (3.13)

$$s\,q(Z^s) - H_s(Z)^A p(Z) = \sum_{j=1}^{s-1} H_s(w^jZ)^A p(w^jZ)\,.$$

Da für $j = 1,\ldots,s-1$ der Punkt $Z = 1$ eine Nullstelle des Polynoms $H_s(w^kZ)$ ist, ist die rechte Seite ein Vielfaches von $(Z-1)^A$. Wegen $H_s(1) = 1$ gibt es ein Inverses $G_s \in \mathbb{Q}[Z]$ zu H_s modulo $(Z-1)^A$. Wir können also zusammenfassend schreiben

$$p(Z) \equiv s\,q(Z)G_s(Z)^A \pmod{(Z-1)^A}\,,$$

d.h. p ist der Rest von $s\,q(Z)G_s(Z)^A$ bei Polynomdivision durch $(1-Z)^A$. Hat q nur rationale Glieder, so hat auch p nur rationale Glieder. □

Korollar 3.5.4 *Seien $s \in \mathbb{N}_{>1}$ und $A, \tilde{A} \in \mathbb{N}_{>0}$. Dann gibt es genau eine Folge $P_{A,\tilde{A}} \in \ell_{\text{fin}}(\mathbb{Q})$ mit Träger im Segment $\{1-\tilde{A},\ldots,A-1\}$ von $\mathbb{Z}$, welche das Gleichungssystem*

$$(\downarrow\text{s})\left(H_s(\mathcal{T})^A H_s(\mathcal{T}^{-1})^{\tilde{A}}\, P_{A,\tilde{A}}\right) = \delta^0$$

erfüllt. Jede andere Lösung dieses Gleichungssystems ist von der Form

$$P = P_{A,\tilde{A}} + (1-\mathcal{T})^A(1-\mathcal{T}^{-1})^{\tilde{A}}R\,,$$

wobei die Folge $R \in \ell_{\text{fin}}(\mathbb{C})$ der Bedingung $(\downarrow\text{s})\,R = 0$ genügen muss.

Beweis: Wir finden nach obigem Satz ein minimales Polynom $p \in \ell_{\text{fin}}(\mathbb{Q})$, welches das Gleichungssystem

$$(\downarrow\text{s})\left(H_s(\mathcal{T})^{A+\tilde{A}}p\right) = \delta^{\tilde{A}}$$

erfüllt. Wegen

$$H_s(Z) = \tfrac{1}{s}(1+Z+\cdots+Z^{s-1}) = Z^{s-1}\tfrac{1}{s}(Z^{1-s}+\cdots+Z^{-1}+1) = Z^{s-1}H_s(Z^{-1})$$

genügt $P_{A,\tilde{A}} := \mathcal{T}^{-\tilde{A}} p$ der Behauptung. Denn

$$\begin{aligned}&\left(\downarrow s\right)\left(H_s(\mathcal{T})^A H_s(\mathcal{T}^{-1})^{\tilde{A}}\, P_{A,\tilde{A}}\right)\\ =&\left(\downarrow s\right)\left(H_s(\mathcal{T})^{A+\tilde{A}}\mathcal{T}^{(1-s)\tilde{A}}\,\mathcal{T}^{-\tilde{A}} p\right)\\ =&\mathcal{T}^{-\tilde{A}}\left(\downarrow s\right)\left(H_s(\mathcal{T})^{A+\tilde{A}}\, p\right) = \delta^0\,.\end{aligned}$$

Das Glied zum Index 0 der Folge $\delta^{\tilde{A}} = \left(\downarrow s\right)\left(H_s(\mathcal{T})^{A+\tilde{A}} p\right)$ muss wegen $\tilde{A} > 0$ einerseits Null sein, andererseits bestimmt es sich als konstanter Koeffizient von $H_s(Z)^{A+\tilde{A}} p(Z)$ zu $s^{-A-\tilde{A}} p_0$, d.h. es gilt $p_0 = 0$. Damit ist der Träger von p auf das Segment $\{1, \ldots, A+\tilde{A}-1\}$ beschränkt. Nach Verschiebung um $-\tilde{A}$ ergibt sich daraus, dass der Träger der Folge $P_{A,\tilde{A}}$ im Segment $\{1-\tilde{A}, \ldots, A-1\} \subset \mathbb{Z}$ enthalten sein muss.

Jede weitere Lösung wird durch eine Folge $R \in \ell_{\text{fin}}(\mathbb{C})$ mit $\left(\downarrow s\right) R = 0$ erzeugt und hat die Gestalt

$$\mathcal{T}^{-\tilde{A}}(p + (1-\mathcal{T})^{A+\tilde{A}} R) = P_{A,\tilde{A}} + (-1)^{\tilde{A}}(1-\mathcal{T})^A (1-\mathcal{T}^{-1})^{\tilde{A}} R\,.$$

□

Beispiel: Seien $s = 3$, $A = \tilde{A} = 2$, und $p, \tilde{p}$ Folgen mit Träger $\{0,1,2\}$, die weiter als symmetrisch vorausgesetzt seien. Mit dem Hilfspolynom

$$X(Z) := \frac{1}{2}(1-Z)(1-Z^{-1}) = 1 - \frac{1}{2}(Z + Z^{-1})$$

gilt

$$H_3(Z) H_3(Z^{-1}) = \frac{1}{9}(Z+1+Z^{-1})^2 = \left(1 - \tfrac{2}{3}X(Z)\right)^2\,,$$

und wir erhalten $P_{2,2}(Z) = \left(1 + \frac{2}{3}X(Z)\right)^4 \mod X(Z)^2$, d.h.

$$P_{2,2}(Z) = 1 + \tfrac{8}{3}X(Z) = \tfrac{4}{3}Z^{-1} + \tfrac{11}{3} - \tfrac{4}{3}Z\,.$$

Mit dem Ansatz

$$\begin{aligned}p(Z) =& Z + \frac{1+\alpha}{4}(1-Z)^2 = Z\left(1 - \frac{1+\alpha}{2}X(Z)\right) \quad \text{und}\\ \tilde{p}(Z) =& Z + \frac{1+\tilde{\alpha}}{4}(1-Z)^2 = Z\left(1 - \frac{1+\tilde{\alpha}}{2}X(Z)\right)\,,\end{aligned}$$

$\alpha, \tilde{\alpha} \in \mathbb{R}$, erhalten wir die reduzierte Biorthogonalitätsbedingung zu

$$p(Z)\tilde{p}^*(Z) = 1 - \tfrac{2+\alpha+\tilde{\alpha}}{2}X(Z) + \tfrac{(1+\alpha)(1+\tilde{\alpha})}{4}X(Z)^2\,,$$

Nach Korollar 3.5.4 müssen $(2+\alpha+\tilde{\alpha}) = -\frac{16}{3}$ und $(1+\alpha)(1+\tilde{\alpha}) = 0$ erfüllt sein, also muss bis auf Vertauschung $\alpha = -1$ und $\tilde{\alpha} = -\frac{16}{3} - 1 = -\frac{19}{3}$ gelten.

Im Beispiel des Abschnitt 6.2.2 (S. 180) wird, der Theorie zur Existenz und Stetigkeit der Skalierungsfunktion folgend, nachgewiesen, dass die Skalierungsfunktion zum Parameter $\alpha = -1$ differenzierbar ist und die Skalierungsfunktion zum Parameter $\alpha = -\frac{16}{3} = -6 - \frac{1}{3}$ stetig ist.

3.5.4 Orthogonale Skalierungsfolgen

Das Ergebnis zur Struktur biorthogonaler Paare von Skalierungsfolgen läßt sich unmittelbar auch für orthogonale Skalierungsfunktionen formulieren.

Korollar 3.5.5 *Es gibt zu $s \in \mathbb{N}_{>1}$ und $A \in \mathbb{N}_{>0}$ ein eindeutig bestimmtes symmetrisches Laurent–Polynom $P_{s,A}(Z) \in \mathbb{C}\langle Z\rangle$ mit Monomträger m–supp $P \subset [1-A, A-1]$, welches die Identität*

$$\big(\downarrow\mathrm{s}\big)\left(H_s(Z)^A H_s(Z^{-1})^A\, P(Z)\right) = \frac{1}{s}$$

erfüllt. Jede andere Lösung dieser Gleichung ist von der Form $P(Z) = P_{s,A}(Z) + \big((Z-1)(Z^{-1}-1)\big)^A R(Z)$ mit einem Laurent–Polynom R, welches $\big(\downarrow\mathrm{s}\big)\, R(Z) = 0$ erfüllt. Ist R symmetrisch, so auch P.

Beweis: Nach dem vorhergehenden Lemma gibt es genau eine Lösung $P_{s,A,A}(Z)$ mit diesem Monomträger. Da $P^*_{s,A,A}(Z) = P_{s,A,A}(Z^{-1})$ ebenfalls die Gleichung löst, muss $P^*_{s,A,A} = P_{s,A,A}$ gelten, $P_{s,A} := P_{s,A,A}$ ist also symmetrisch. □

Um aus einer Lösung $P \in \ell_{\mathrm{fin}}(\mathbb{R})$ dieses linearen Gleichungssystems eine orthogonale Skalierungsfolge p zu erhalten, muss das Laurent–Polynom $P(Z)$ faktorisiert werden. Beide Faktoren ergeben sich aus der Folge p, insbesondere gilt für die trigonometrischen Polynome die Identität $\hat{P}(\omega) = |\hat{p}(\omega)|^2$. Es muss also geprüft werden, ob $\hat{P}$ nur nichtnegative Werte annimmt und ob sich daraus eine Faktorisierung ergibt.

Lemma 3.5.6 (s. [Hel95]) *Für jedes $s \in \mathbb{N}_{>1}$ und $A \in \mathbb{N}_{>0}$ gibt es ein Laurent–Polynom $p(Z)$ mit reellen Koeffizienten, so dass $p(Z)p^*(Z) = P_{s,A}(Z)$ gilt. Dabei ist $P_{s,A}(Z)$ das symmetrische Laurent–Polynom, welches im vorangegangenen Korollar erhalten wurde.*

Wir erinnern daran, dass bei reellen Koeffizienten nach Definition $p^*(Z) = p(Z^{-1})$ gilt. Weiterhin ist $P_{s,A}(Z)$ nach dem vorangegangenen Korollar das kleinste symmetrische Laurent–Polynom, welches

$$P_{s,A}(Z) H_s(Z)^A H_s^*(Z) \equiv 1 \mod (1-Z)^A (1-Z^{-1})^A$$

erfüllt.

Beweis: Diese Aussage ist eine Folgerung aus dem Fejer–Riesz–Algorithmus (s. [PS71], nach [BKN94]): Gegeben sei ein symmetrisches Laurent–Polynom $P \in \mathbb{R}\langle Z\rangle$, $P(Z) = P^*(Z)$ mit reellen Koeffizienten. Hat die Auswertung $P(z)$, $z \in \mathbb{C}$, auf dem Einheitskreis der komplexen Ebene, d.h. bei $|z| = 1$, nur reelle, nichtnegative Werte, so gibt es eine Faktorisierung dieser Art.

Nach Lemma 3.4.6 gibt es zum symmetrischen Laurent–Polynom $P_{s,A}(Z)$ ein Polynom $Q_{s,A} \in \mathbb{Q}[X]$ mit $P_{s,A}(e^{i\omega}) = Q_{s,A}(1-\cos(\omega))$. Kann nachgewiesen werden, dass alle Koeffizienten von $Q_{s,A}$ positiv sind, so nimmt auch $P_{s,A}(Z)$ auf dem Einheitskreis nur positive reelle Werte an und kann somit mittels des Fejer–Riesz–Algorithmus faktorisiert werden.

Für das symmetrische Laurent–Polynom $X := \frac{1}{2}(1-Z)(1-Z^{-1}) = 1 - \frac{1}{2}(Z+Z^{-1})$ gilt $X(e^{i\omega}) = 1-\cos(\omega)$. Sei $p \in \mathbb{R}[Z]$ ein beliebiges Polynom des Grades $d := \deg p$ mit reellen Koeffizienten und Nullstellen $z_1, \ldots, z_d$. Dann gilt

$$\begin{aligned} p(Z)p(Z^{-1}) &= \prod_{k=1}^{d}(Z-z_k)(Z^{-1}-z_k) = \prod_{k=1}^{d}\left(1 - z_k(Z+Z^{-1}) + z_k^2\right) \\ &= \prod_{k=1}^{d} 2z_k\left(X - 1 + \tfrac{1}{2}(z_k + z_k^{-1})\right) = (-2)^d p_0 \prod_{k=1}^{d}(X - X(z_k)) \end{aligned}$$

Im letzten Schritt wurde der Vietasche Wurzelsatz benutzt. Im Fall des Haar–Polynoms gilt $d = s-1$, $p_0 = \frac{1}{s}$ und $z_k = e^{i(2\pi s^{-1})k}$, $k = 1, \ldots, s-1$. Daher folgt $X(z_k) = 1 - \cos(2\pi s^{-1}k) = 2\sin^2(\pi s^{-1}k)$ und weiter

$$H_s(Z)H_s(Z^{-1}) = \frac{2^{s-1}}{s}\prod_{k=1}^{s-1}(2\sin^2(\pi s^{-1}k) - X)\,.$$

An der Stelle $Z = 1$ ausgewertet ergibt sich daraus mit $X(1) = 0$ und $H_s(1) = 1$

$$1 = \frac{4^{s-1}}{s}\prod_{k=1}^{s-1}\sin^2(\pi s^{-1}k)$$

und daraus wiederum

$$H_s(Z)H_s(Z^{-1}) = \prod_{k=1}^{s-1}\left(1 - \frac{X}{2\sin^2(\pi s^{-1}k)}\right)\,.$$

Die rechte Seite kann nun einfach modulo X^A invertiert werden, da für beliebige $c \in \mathbb{R}$

$$(1-cX)(1+cX+c^2X^2+\cdots+c^{A-1}X^{A-1}) = 1 - c^A X^A$$

gilt. Die Koeffizienten des Polynoms $Q_{s,A}$ können somit als die Koeffizienten bis zum Grad $A-1$ des Ausdrucks

$$\prod_{k=1}^{s-1}\sum_{m=0}^{A-1}\left(\frac{X}{2\sin^2(\pi s^{-1}k)}\right)^m$$

abgelesen werden. Dieses Produkt ist aus Faktoren zusammengesetzt, welche sämtlich nur positive Koeffizienten haben. Dies bleibt beim Ausmultiplizieren erhalten, somit hat auch $Q_{s,A}(X)$ nur positive Koeffizienten. Insbesondere gilt $Q_{s,A}(0) = 1$. □

Für praktische Rechnungen ist es vorteilhafter, statt der eben benutzten Faktorisierung das Produkt $H_s(Z)H_s(Z^{-1})$ direkt als Polynom in X mit rationalen Koeffizienten darzustellen. Eine solche Darstellung kann man mit der Methode der erzeugenden formalen Potenzreihen ableiten. Sei t ein formaler Parameter, dann gilt

$$\begin{aligned} \sum_{s=1}^{\infty} t^{s-1}s^2 H_s(Z)H_s^*(Z) &= \sum_{0\le k<s} t^{s-1}(s-k)(Z^k+Z^{-k}) - \sum_{s=1}^{\infty} t^{s-1}s \\ &= \sum_{n=0}^{\infty} t^{n-1}n\sum_{k=0}^{\infty}\left((tZ)^k + (tZ^{-1})^k\right) - \tfrac{1}{(1-t)^2} \\ &= \tfrac{1}{(1-t)^2}\left(\tfrac{1}{1-tZ} + \tfrac{1}{1-tZ^{-1}} - 1\right) = \tfrac{1}{(1-t)^2}\left(\tfrac{1-t^2}{1-t(Z+Z^{-1})+t^2}\right) \end{aligned}$$

An dieser Stelle kann $Z + Z^{-1} = 2(1 - X)$ eingesetzt werden und der rationale Ausdruck in eine geometrischen Reihe entwickelt werden,

$$\begin{aligned}
&\tfrac{1+t}{1-t}\ \tfrac{1}{(1-t)^2+2tX} \\
&\qquad = \sum_{k=0}^{\infty} \tfrac{1+t}{(1-t)^{2k+3}}(-2t)^k X^k = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \tbinom{2k+2+n}{2k+2}(1+t)t^{k+n}(-2X)^k \\
&\qquad = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \tbinom{2k+1+n}{2k+1}\tfrac{2k+2+n+n}{2k+2} t^{k+n}(-2X)^k
\end{aligned}$$

Durch Vergleich der Koeffizienten des formalen Parameters t ergibt sich schließlich

$$\begin{aligned}
H_s(Z)H_s^*(Z) =& \frac{1}{s^2}\sum_{k=0}^{s-1} \tbinom{s+k}{2k+1}\ \tfrac{s}{k+1}(-2X)^k = \sum_{k=0}^{s-1} \tfrac{\prod_{n=1}^{k}(s^2-n^2)}{(k+1)(2k+1)!}(-2X)^k \\
=& 1 - \tfrac{s^2-1}{6}X + \tfrac{(s^2-1)(s^2-4)}{90}X^2 \pm \dots
\end{aligned} \tag{3.14}$$

Beispiel: Für die polynomiale Approximationsordnung $A = 1$ gilt $P_{s,1}(Z) = 1$, was vom Skalenfaktor s unabhängig ist. Für die polynomiale Approximationsordnung $A = 2$ ergibt sich das Polynom $Q_{s,2}$ aus der Identität

$$\left(1 - \tfrac{s^2-1}{6}X\right)^2 Q_{s,2}(X) \equiv 1 \bmod X^2$$

zu $Q_{s,2}(X) = 1 + \frac{(s^2-1)}{3}X$. Aus dem Ansatz $p(Z) = \frac{1}{2}(1 + Z + \kappa(1 - Z))$ ergibt sich

$$P(Z) = p(Z)p(Z^{-1}) = \frac{1}{4}(4 - 2X + \kappa^2 2X) = 1 + \frac{\kappa^2 - 1}{2}X\,.$$

Soll die rechte Seite mit $Q_{s,2}$ übereinstimmen, so muss $\kappa^2 - 1 = 2\frac{s^2-1}{3}$, d.h. $\kappa = \sqrt{\frac{2s^2+1}{3}}$ gelten.

Bei $s = 2$ erhält man damit $\kappa = \sqrt{3}$, und daraus eine Skalierungsfolge a als Koeffizientenfolge des Laurent–Polynoms

$$a(Z) = \frac{1}{4}\left((1+Z)^3 + \alpha(1+Z)(1-Z^2)\right) = \frac{1+\sqrt{3}}{4} + \frac{3+\sqrt{3}}{4}Z + \frac{3-\sqrt{3}}{4}Z^2 + \frac{1+\sqrt{3}}{4}Z^3\,.$$

Diese Skalierungsfolge ergibt das als D4 bezeichnete einfachste Daubechies–Wavelet. Die Abschätzung der Stetigkeit der Skalierungsfunktion ergibt sich – unter Vorgriff auf Kapitel 6 – aus der Größe der Norm

$$\|\,(\downarrow 2)\,p\|_{(1,\infty)} = \frac{1}{2}\max|1+\kappa|, |1-\kappa| = \frac{1}{2}(1+\sqrt{3}) \le 2^{0.45} = 2^{2-1-0.55}\,.$$

Die Skalierungsfunktion zu dieser Skalierungsfolge ist daher Hölder–stetig mit Index 0.55, und die Wavelet–Funktion erzeugt eine orthogonale Wavelet–Basis.

Bei $s = 3$ erhalten wir $\kappa = \sqrt{\frac{19}{3}}$ mit $\frac{1}{2}(1 + \sqrt{6 + \frac{1}{3}}) \le 3^{2-1-0.48}$, bei $s = 4$ analog $\kappa = \sqrt{11}$ mit $\frac{1}{2}(1 + \sqrt{11}) \le 4^{2-1-0.44}$ usw.

3.5.5 Optimierungsproblem für orthogonale Skalierungsfunktionen

Es sei ein $s \in \mathbb{N}_{>1}$ und eine polynomiale Approximationsordnung $A \in \mathbb{N}_{>0}$ fixiert, ebenso ein Segment $\{M, \dots, N\} \subset \mathbb{Z}$, welches als Träger einer endlichen Folge $p \in \ell_{\text{fin}}(\mathbb{R})$ dienen soll.

Unter allen endlichen Folgen $p \in \ell_{\text{fin}}(\mathbb{R})$ mit $\operatorname{supp} p \subset \{M, \dots, N\}$, für deren Laurent–Polynom $p(1) = \sqrt{s}$ gilt, betrachten wir diejenigen, für welche es eine ebenfalls endliche Folge $R \in \ell_{\text{fin}}(\mathbb{R})$ mit $\operatorname{supp} R \subset \{M - N + A, \dots, N - M - A\}$ gibt, so dass die Identitäten

- $$\big(\downarrow\mathrm{s}\big)\, R = 0 \quad \text{und} \tag{3.15a}$$
- $$p(Z)p(Z^{-1}) = P_A(Z) + \Big((1-Z)(1-Z^{-1})\Big)^A\, R(Z) \tag{3.15b}$$

erfüllt sind.

Auf der Menge dieser Folgen werden nun diejenigen gesucht, deren trigonometrisches Polynom möglichst kleine Werte hat. Ein – recht grobes – Maß dafür ist die Quadratsumme $\sum_{k=M}^{N} (p_k)^2$ der Koeffizienten der Folge.

3.5.6 Symmetrische orthogonale Skalierungsfunktionen

Sei ein $s \in \mathbb{N}_{>1}$ fixiert. Alle bisher untersuchten Eigenschaften zusammenfassend betrachten wir hier symmetrische orthogonale Skalierungsfunktionen $a \in \ell_{\text{fin}}(\mathbb{R})$, deren Tiefpassfilter $a(\mathcal{T})\,\big(\uparrow\mathrm{s}\big)$ die Symmetrie mit Verzögerung $d = -1$ erhält.

Nach Satz 3.3.3 zu symmetrieerhaltenden Filterbänken muss dann a symmetrisch mit Verzögerung $s-1$ sein, d.h. es gilt

$$a(Z) = \sum_{n\in\mathbb{Z}} a_n Z^n = \sum_{n\in\mathbb{Z}} a_{s-1-n} Z^n = Z^{s-1} a(Z^{-1})\,.$$

Das Haar–Polynom hat diese Art der Symmetrie. Für eine beliebige Skalierungsfolge $a(Z) = sH_s(Z)^A p(Z)$ mit $a \in \mathbb{N}_{>0}$ und einer endlichen Folge $p \in \ell_{\text{fin}}(\mathbb{R})$ folgt aus der Symmetrie von a die Identität

$$sH_s(Z)^A p(Z) = sZ^{s-1} H_s(Z^{-1})^A p(Z^{-1}) = sZ^{(1-A)(s-1)} H_s(Z)^A p(Z^{-1})\,,$$

die Folge p muss also symmetrisch mit Verzögerung $(1-A)(s-1)$ sein. Ist diese Verzögerung gerade, so entsteht p durch Verschiebung aus einer zum Index Null symmetrischen Folge,

$$p(Z) = Z^M q(X) \quad \text{bei} \quad 2M = (1-A)(s-1)\,.$$

Dabei ist $q \in \mathbb{R}[X]$ ein Polynom, $X = 1 - \frac{1}{2}(Z + Z^{-1})$ ist dasselbe minimale symmetrische Laurent–Polynom wie in Lemma 3.4.6.

Ist die Verzögerung ungerade, d.h. sowohl A als auch s sind gerade, so enthält das Laurent–Polynom $p(Z)$ einen Faktor $(1+Z)$ und einen weiteren symmetrischen Faktor gerader Verzögerung, kann also als

$$p(Z) = \frac{1}{2}(1+Z)Z^M q(X) \quad \text{bei} \quad 2M+1 = (1-A)(s-1)$$

und $q \in \mathbb{R}[X]$ dargestellt werden.

Satz 3.5.7 *Seien $s \in \mathbb{N}_{>1}$, $A \in \mathbb{N}_{>0}$ und $C_{s,A} \in \mathbb{Q}[X]$ mit $C_{s,A}(X) = 1$ wenn eine der Zahlen s und A ungerade ist, $C_{s,A}(X) = 1 - \frac{1}{2}X$ im gegenteiligen Fall, wenn beide gerade sind.*

Jede endliche symmetrische orthogonale Skalierungsfolge $a \in \ell_{\text{fin}}(\mathbb{R})$ zum Skalenfaktor s und mit polynomialer Approximationsordnung A ergibt sich als Lösung des Problems, Polynome $q, r \in \mathbb{R}[X]$ zu finden, deren Koeffizienten das aus der Identität

$$C_{s,A}(x)q(X)^2 = Q_{s,A}(X) + X^A\, r(X)$$

resultierende polynomiale Gleichungssystem erfüllen. Zusätzlich müssen für die Koeffizientenfolge des Laurent–Polynoms $R(Z) := r\big(1 - \frac{1}{2}(Z + Z^{-1})\big)$ die aus der Identität $0 = \big(\downarrow\mathrm{s}\big)\, R$ abgeleiteten linearen Gleichungen erfüllt sein.

Beweis: Diese Aussage ergibt sich aus der eben abgeleiteten Struktur symmetrischer Skalierungsfolgen, kombiniert mit Korollar 3.5.5 zu den Eigenschaften orthogonaler Skalierungsfolgen. Das Hilfspolynom $C_{s,A}$ ergibt sich für gerades s und A aus

$$\frac{1}{4}(1+Z)(1+Z^{-1}) = \frac{1}{4}\left(2 + Z + Z^{-1}\right) = 1 - \frac{1}{2}\left(1 - \frac{1}{2}(Z + Z^{-1})\right)\ .$$

□

Die eben angegebene polynomiale Identität für die Polynome q und r kann modulo X^A direkt, d.h. mit rein arithmetischen Mitteln, gelöst werden. Denn es gilt $P_{s,A}(Z) = Q_{s,A}(X(Z))$, und $Q_{s,A}$ ist ein Polynom mit konstantem Koeffizienten 1.

Ist $s = 2m + 1$ ungerade mit $m \in \mathbb{N}_{>0}$, so ist $Z^{-m}H_s(Z)$ symmetrisch zum Index 0, kann also mit einem Polynom $h_s \in \mathbb{Q}[X]$ als $Z^{-m}H_s(Z) = h_s(X)$ dargestellt werden. Das Polynom $Q_{s,A}$ ist dann die minimale Lösung der Gleichung

$$(H_s(Z)H_s(Z^{-1}))^A P_{s,A} = h_s(X)^{2A} Q_{s,A}(X) \equiv 1 \bmod X^A\ .$$

Die Wurzel aus $Q_{s,A}$ und damit die ersten A Koeffizienten von q ergeben sich daher aus der Gleichung

$$q(X) \equiv h_s(X)^{-A} \bmod X^A\ .$$

Nach erfolgter Bestimmung der übrigen Koeffizienten von q ergibt sich die Skalierungsfolge als

$$a(Z) = sZ^m h_s(X)^A q(X)\ .$$

Ist $s = 2m$ gerade mit $m \in \mathbb{N}_{>0}$, so enthält $H_s(Z)$ einen Faktor $1 + Z$, es gilt

$$H_s(Z) = Z^{m-1}\frac{1}{2}(1+Z)\ \frac{1}{m}(Z^{1-m} + Z^{3-m} + \cdots + Z^{m-1})\ .$$

Der zweite symmetrische Faktor kann wieder als Polynom $h_s \in \mathbb{Q}[X]$ im symmetrischen Baustein X ausgedrückt werden. Somit gilt

$$H_s(Z)H_s(Z^{-1}) = (1 - \tfrac{1}{2}X)h_s(X)^2\ .$$

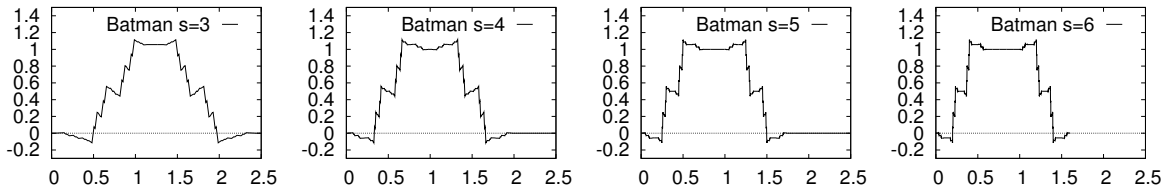


Abbildung 3.1: Batman–Wavelets, Skalierungsfunktionen für Faktor $s = 3, \ldots, 6$

Ist nun A gerade, $A = 2B$, $B \in \mathbb{N}_{>0}$, so ergeben sich die Koeffizienten der niedrigen Grade von q aus

$$(1 - \tfrac{1}{2}X)q(X)^2 \equiv Q_{s,A}(X) \bmod X^A$$

und daher

$$q(X) \equiv h_s(X)^{-A}(1 - \tfrac{1}{2}X)^{-B-\frac{1}{2}} \bmod X^A \; .$$

Dieselbe Formel ergibt sich auch für ungerades $A = 2B + 1$, $B \in \mathbb{N}_0$. Nach erfolgter Bestimmung der übrigen Koeffizienten von q ergibt sich die Skalierungsfolge in beiden Fällen als

$$a(Z) = mZ^m(1 + Z)(1 - \tfrac{1}{2}X)^B h_s(X)^A q(X) \; .$$

Beispiel: In [BW99] wurde die Klasse der *„Batman"–Skalierungsfunktionen* eingeführt, welche für beliebigen Skalenfaktor $s \in \mathbb{N}$, $s \geq 3$ und Approximationsordnung $A = 1$ die symmetrischen stetigen Lösungen geringster Breite des Monom–Trägers darstellen. Dazu wurde der Ansatz

$$p(Z) := \tfrac{1+Z}{2}(Z + 2\alpha(1 - Z)^2) = \alpha + (\tfrac{1}{2} - \alpha)Z + (\tfrac{1}{2} - \alpha)Z^2 + \alpha Z^3$$

gemacht. Da $A = 1$ ungerade ist, aber p nicht symmetrisch zum Index Null, widerspricht dieser Ansatz der oben formulierten anwendungsbezogenen Symmetriebedingung. Sei wieder $X(Z) := \frac{1}{2}(1 - Z)(1 - Z^{-1})$, dann ist $p(Z) = \frac{1}{2}Z(1 + Z)(1 - 4\alpha X(Z))$.

In der Zerlegung $p(Z)p^*(Z) = 1 + X(Z)R(Z)$ muss $\big(\downarrow\mathrm{s}\big)\, R(Z) = 0$ gelten. Nach Ausmultiplizieren erhält man

$$\begin{aligned} p(Z)p^*(Z) =& \Big(1 - \tfrac{1}{2}X(Z)\Big)\Big((1 - 8\alpha X(Z) + 16\alpha^2 X(Z)^2\Big) \\ =& 1 - X(Z)\ \Big(\tfrac{1}{2} + 8\alpha - 4(\alpha + 4\alpha^2)X(Z) + 8\alpha^2 X(Z)^2\Big) \\ 0 = \big(\downarrow\mathrm{s}\big)\, R(Z) =& \tfrac{1}{2} + 8\alpha - 4(\alpha + 4\alpha^2) + 12\alpha^2 = \tfrac{1}{2} + 4\alpha - 4\alpha^2 = \tfrac{3}{2} - (1 - 2\alpha)^2 \\ =& (2\alpha - 1 - \tfrac{\sqrt{6}}{2})(2\alpha - 1 + \tfrac{\sqrt{6}}{2}) \end{aligned}$$

Die Kenngröße zur Bestimmung der Existenz stetiger Lösungen der Verfeinerungsgleichung lautet für $s = 3$

$$\| \big(\downarrow\downarrow 3\big)\, p\|_{(1,\infty)} = \max(2|\alpha|, |\tfrac{1}{2} - \alpha|) = \begin{cases} 1 + \frac{\sqrt{6}}{2} > 3^{0{,}7278657} & \text{für } \alpha = \frac{1}{2} + \frac{\sqrt{6}}{4} \\ \frac{\sqrt{6}}{4} < 3^{-0{,}4463946} & \text{für } \alpha = \frac{1}{2} - \frac{\sqrt{6}}{4} \end{cases} ,$$

d.h. keine Lösung nach Satz 6.2.6 im ersten und eine stetige Lösung mit Hölder–Index $\alpha = 0.446\ldots$ im zweiten Fall. Für $s > 3$ erhalten wir

$$\| \big(\downarrow\mathrm{s}\big)\, p\|_{(1,\infty)} = \max(|\alpha|, |\frac{1}{2} - \alpha|) = \begin{cases} \frac{1}{2} + \frac{\sqrt{6}}{4} \approx 1.11237 & \text{für } \alpha = \frac{1}{2} + \frac{\sqrt{6}}{4} \\ \frac{\sqrt{6}}{4} < \frac{3}{4} & \text{für } \alpha = \frac{1}{2} - \frac{\sqrt{6}}{4} \end{cases} ,$$

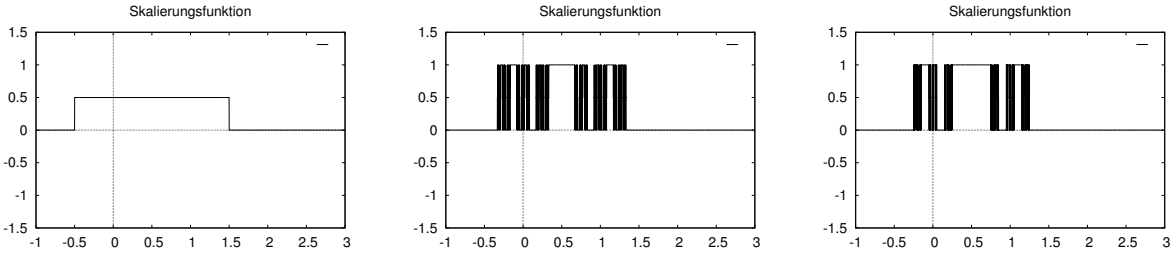


Abbildung 3.2: Approximationen der Skalierungsfunktionen für $A = 1$, eine Variable und $s = 3, 4, 5$

d.h. im ersten Fall erhalten wir mit Satz 6.2.6 keine Lösung, die Stetigkeit der zweiten Lösung bleibt erhalten. Nach Satz 6.2.2 erhalten wir auch im ersten Fall für $s \geq 4$ Lösungen in $L^1(\mathbb{R})$ und für $s \geq 12$ Lösungen in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, denn es ist

$$\|p\|_{\ell_1} = 2|\alpha| + 2|\tfrac{1}{2} - \alpha| \approx 3.44949 \approx \sqrt{11.89898}\,.$$

3.5.7 Einfache Beispiele symmetrischer orthogonaler Skalierungsfunktionen

In dem Ansatz zur Berechnung symmetrischer orthogonaler Skalierungsfunktionen stehen die Anzahl der Variablen zur Anzahl der aus den Orthogonalitätsbedingungen gewonnenen quadratischen Gleichungen im Verhältnis von etwa $s : 1$ für einen gegebenen Skalierungsfaktor s. Um eine aus isolierten Punkten bestehende Lösungsmenge zu erhalten, dürfen also nur eine oder zwei Variablen im Ansatz verwendet werden. Es stellt sich heraus, dass es nur endlich viele Kombinationen der von Skalenfaktor $s \in \mathbb{N}_{\geq 2}$ und Approximationsordnung $A \in \mathbb{N}_{\geq 1}$ gibt, für welche dieser Ansatz reelle Lösungen mit stetigen Skalierungsfunktionen besitzt. Einige dieser Beispiele und weitere Beispiele, welche die Symmetrie nicht in der hier geforderten Weise berücksichtigen, finden sich in [BW99] und [Han98].

Sei zunächst die Approximationsordnung $A = 1$ bei beliebigem Skalenfaktor $s \in \mathbb{N}_{>1}$ betrachtet. Die Symmetrie der Skalierungsfolge soll den oben abgeleiteten Anforderungen genügen, also Verzögerung $(s-1)$ aufweisen. Die Skalierungsfolge hat also die Form $a(Z) = sH_s(Z)q(X)$. Ein erster Versuch mit einen zusätzlichen Parameter $c \in \mathbb{R}$, d.h. $q(X) := 1 + cX$, führt auf

$$q(X)^2 = 1 + X(2c + c^2X), \implies r(X) = c\,(2 + cX) \implies R(Z) = c(-\frac{c}{2}Z^{-1} + (2+c) - \frac{c}{2}Z)\,.$$

Die Bedingung $\big(\downarrow\mathrm{s}\big)\, R = 0$ ist somit stets für $c = 0$ und $c = -2$ erfüllt. Für $c = 0$ ergibt sich das Haar–Polynom ohne zusätzlichen Faktor, für $c = -2$ ergibt sich der zusätzliche Faktor $q(X) = 1 - 2X = Z^{-1} - 1 + Z = p(Z)$. Für keinen Skalierungsfaktor ergibt sich aus dieser Skalierungsfolge eine stetige Skalierungsfunktion (s. Abbildung 3.2). Durch das Hinzunehmen weiterer Parameter ergeben sich freie Parameter, so dass ein polynomiales Optimierungsproblem aufzustellen und zu lösen ist, um reelle Lösungen zu bestimmen.

Betrachten wir den Fall $A = 2$ mit $Q_{s,2}(X) = 1 + \frac{s^2-1}{3}X$ und einem Ansatz $q(X) = 1 + c_1X + c_2X^2$. Dabei ist c_1 eine von s abhängige Konstante, c_2 eine Variable. Modulo X^2 ergibt sich für ungerades s

$$q^2 \equiv 1 + 2c_1X \equiv 1 + \frac{s^2-1}{3}X \bmod X^2\,,$$

d.h. $c_1 = \frac{s^2-1}{6}$. Für gerades s muss

$$(1 - \tfrac{1}{2}X)q^2 \equiv 1 + 2c_1X - \tfrac{1}{2}X \equiv 1 + \frac{s^2-1}{3}X \bmod X^2$$

gelten, d.h. $c_1 = \frac{s^2-1}{6} + \frac{1}{4}$. Der verbleibende Koeffizient c_2 bestimmt sich aus der quadratischen Gleichung $R_0 = 0$, wobei das Laurent–Polynom $R(Z)$ sich aus der Identität $R(Z) = r(X)$ ergibt, und das Polynom r aus der Identität

$$X^A r(X) = C_{s,2}(X)q(X)^2 - Q_{s,2}(X)$$

erhalten wird. Für ungerades s ergibt sich, nach Zusammenfassen der Terme im Sinne einer quadratischen Ergänzung, die Gleichung

$$0 = R_0 = \tfrac{1}{3}(c_1 - 2)^2 + \tfrac{3}{2}(c_2 + \tfrac{2}{3}(c_1 + 1))^2 - 2\,.$$

Für $s \geq 7$ wird der erste Summand schon größer als 2, wodurch nur für $s = 3$ und $s = 5$ reelle Lösungen für c_2 möglich sind. Für geraden Skalenfaktor s ergibt sich analog

$$0 = R_0 = \tfrac{1}{4}(c_1 - 4)^2 + \tfrac{1}{4}(c_2 + c_1 + 2)^2 - 5\,.$$

Auch hier ist für $s \geq 8$ die Differenz aus dem ersten Summanden und der Konstanten 5 positiv, wodurch nur für $s = 2$, $s = 4$ und $s = 6$ reelle Lösungen möglich sind. Der Grad von $r(X)$ beträgt 3, im Fall $s = 2$ ergibt sich aus der Orthogonalitätsbedingung $\big(\downarrow 2\big)\, R = 0$ eine weitere nichttriviale Gleichung, nämlich $R_2 = 0$. Diese widerspricht jedoch der ersten Gleichung, so dass dieser Fall keine Lösungen besitzt. In allen weiteren Fällen ergeben sich keine weiteren Gleichungen für c_2. Es ergeben sich je zwei reelle Lösungen. Diese sind im Folgenden zusammengestellt.

Für $s = 3$ ergibt sich $c_1 = \frac{3}{4}$ und damit $c_2 \in \{-\frac{8}{3}, -\frac{4}{9}\}$. Die Skalierungsfolgen sind

$$a(Z) = 3Z\left(\frac{Z + 1 + Z^{-1}}{3}\right)^2 (-\tfrac{2}{3}Z^2 + \tfrac{55}{24}Z - \tfrac{9}{4} + \tfrac{55}{24}Z^{-1} - \tfrac{2}{3}Z^{-2})$$

mit einer stetigen Skalierungsfunktion mit Hölder–Index 0.12 (s. Abb. 3.3 links) und

$$a(Z) = 3Z\left(\frac{Z + 1 + Z^{-1}}{3}\right)^2 (-\tfrac{1}{9}Z^2 + \tfrac{5}{72}Z + \tfrac{13}{12} + \tfrac{5}{72}Z^{-1} - \tfrac{1}{9}Z^{-2})$$

mit einer stetigen Skalierungsfunktion mit Hölder–Index 0.92 (s. Abb. 3.3 rechts).

Für $s = 4$ ergibt sich $c_1 = \frac{11}{4}$ und damit $c_2 = -\frac{19}{4} \pm \frac{\sqrt{295}}{4}$. Die resultierenden Skalierungsfolgen sind

$$a(Z) \approx 2(1 + Z)\left(\frac{Z + 1 + Z^{-1} + Z^2}{4}\right)^2 \cdot \ldots$$
$$\ldots \cdot (-2.26097\, Z^2 + 7.66889\, Z - 9.81584 + 7.66889\, Z^{-1} - 2.26097\, Z^{-2})$$

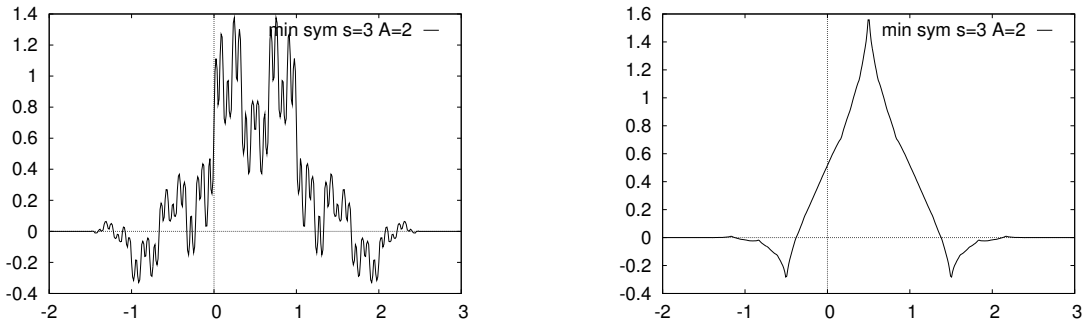


Abbildung 3.3: symmetrische Skalierungsfunktionen minimaler Breite für $s = 3$, $A = 2$

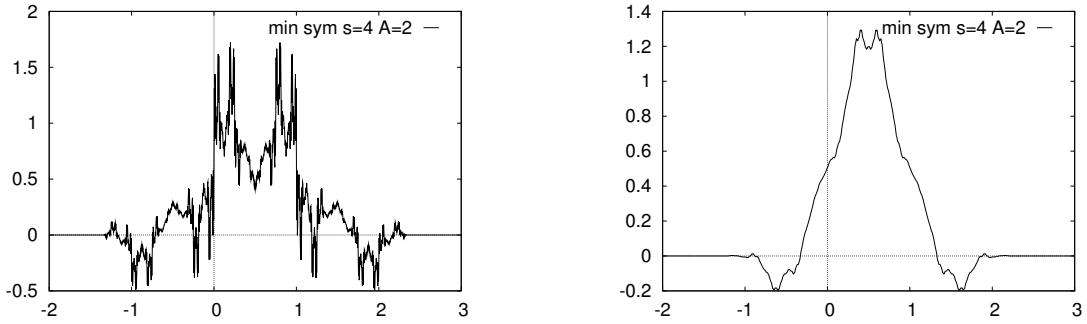


Abbildung 3.4: symmetrische Skalierungsfunktionen minimaler Breite für $s = 4$, $A = 2$

mit einer stetigen Skalierungsfunktion mit Hölder–Index 0.03 (s. Abb. 3.4 links) und

$$a(Z) \approx 2(1+Z)\left(\frac{Z^2+Z+1+Z^{-1}}{4}\right)^2 \cdot \ldots$$
$$\ldots \cdot (-0.11403\,Z^2 - 0.91889\,Z + 3.06584 - 0.91889\,Z^{-1} - 0.11403\,Z^{-2})$$

mit einer stetigen Skalierungsfunktion mit Hölder–Index 0.94 (s. Abb. 3.4 rechts).

Für $s = 5$ ergibt sich $c_1 = 4$ und damit $c_2 \in \{-4, -\frac{8}{3}\}$. Die Skalierungsfolgen sind

$$a(Z) \quad = \quad 5Z^2\left(\frac{Z^2+Z^1+1+Z^{-1}+Z^{-2}}{5}\right)^2 \cdot (-Z^2 + 2Z^1 - 1 + 2Z^{-1} - Z^{-2})$$

mit einer stetigen Skalierungsfunktion mit Hölder–Index 0.56 (s. Abb. 3.5 links) und

$$a(Z) \quad = \quad 5Z^2\left(\frac{Z^2+Z^1+1+Z^{-1}+Z^{-2}}{5}\right)^2 \cdot (-\tfrac{2}{3}Z^2 + \tfrac{2}{3}Z^1 + 1 + \tfrac{2}{3}Z^{-1} - \tfrac{2}{3}Z^{-2})$$

mit einer stetigen Skalierungsfunktion mit Hölder–Index 0.99 (s. Abb. 3.5 rechts).

Für $s = 6$ ergibt sich $c_1 = \frac{73}{12}$ und damit $c_2 = -\frac{97}{12} \pm \frac{\sqrt{2255}}{12}$. Die Skalierungsfolgen sind

$$a(Z) \approx 3Z(1+Z)\left(\frac{Z^3+Z^2+Z+1+Z^{-1}+Z^{-2}}{6}\right)^2 \cdot \ldots$$
$$\ldots \cdot (-3.01014\,Z^2 + 8.99890\,Z - 10.97752 + 8.99890\,Z^{-1} - 3.01014\,Z^{-2})$$

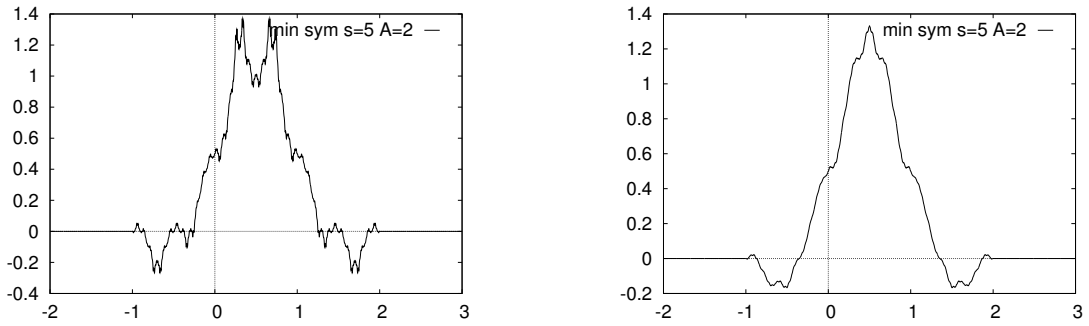


Abbildung 3.5: symmetrische Skalierungsfunktionen minimaler Breite für $s = 5$, $A = 2$

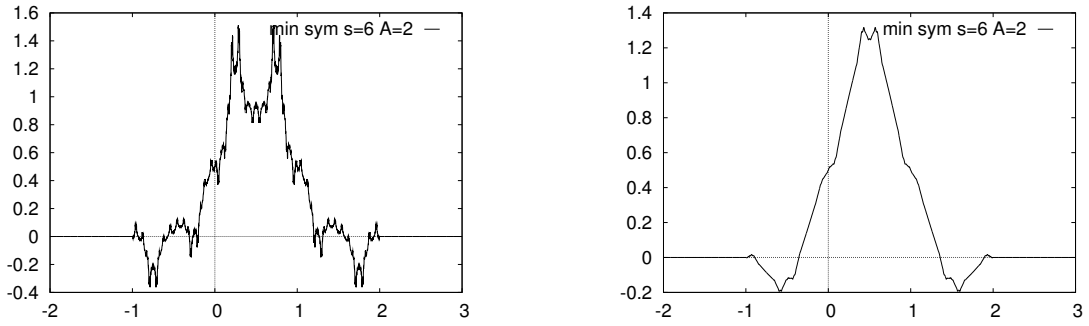


Abbildung 3.6: symmetrische Skalierungsfunktionen minimaler Breite für $s = 6$, $A = 2$

mit einer stetigen Skalierungsfunktion mit Hölder–Index 0.38 (s. Abb. 3.6 links) und

$$a(Z) \approx 3Z(1+Z)\left(\frac{Z^3+Z^2+Z+1+Z^{-1}+Z^{-2}}{6}\right)^2 \cdot \ldots$$
$$\ldots \cdot (-1.03152\,Z^2 + 1.08443\,Z + 0.89419 + 1.08443\,Z^{-1} - 1.03152\,Z^{-2})$$

mit einer stetig differenzierbaren Skalierungsfunktion mit Hölder–Index 0.005 der ersten Ableitung (s. Abb. 3.6 rechts).

Im Falle einer Approximationsordnung 3 ist immer $C_{s,3}(X) = 1$, es ergibt sich aus der arithmetischen Auswertung der Orthogonalitätsbedingungen

$$Q_{s,3}(X) = 1 + \tfrac{1}{2}(s^2-1)\,X + \tfrac{1}{30}(s^2-1)(4s^2-1)\,X^2$$

und

$$q_{s,3}(X) = 1 + \tfrac{1}{4}(s^2-1)\,X + \tfrac{1}{480}(s^2-1)(17s^2+7)\,X^2\,.$$

Um den Faktor q der Skalierungsfolge zu erhalten, muss $q_{s,3}$ um weitere Terme der Ordnung 3 und höher erweitert werden. Im hier betrachteten einfachsten Fall ist $q(X) = q_{s,3}(X) + c\,X^3$. Die Auswertung der verbleibenden Orthogonalitätsbedingungen ergibt ein quadratisches Polynom in c mit Parameter s, welches für $s \geq 4$ und reelles c immer positiv ist. Für $s = 3$ ergibt sich zusätzlich die Bedingung $c^2 = 0$, was aber mit der ersten Bedingung in Widerspruch steht. Es gibt daher für keinen Skalenfaktor s eine reelle Lösung. Es ist zu vermuten, dass sich dieses Verhalten für alle weiteren Approximationsordnungen $A \geq 4$ fortsetzt.

3.6 Zur Vervollständigung von Waveletfilterbänken

3.6.1 Faktorisierung von semi–unitärer Differenzenoperatoren

Angenommen, obiges System quadratischer Gleichungen ist gelöst und eine Folge $a \in \ell_{\text{fin}}(\mathbb{R})$ liegt vor, für die der Operator $a(\mathcal{T})\,(\uparrow s)$ semi–unitär ist. Um eine orthogonale Wavelet–Filterbank zu erhalten, muss die einspaltige Polyphasenmatrix dieses $(1,s)$–periodischen Operators um weitere $(s-1)$ Spalten ergänzt werden, so dass die volle Matrix unitär ist.

Es ist auf einfache Weise möglich, von semi–unitären Differenzenoperatoren unitäre Faktoren mit zweigliedrigen Koeffizientenfolgen abzuspalten und auf diese Weise den semi–unitären Operator auf einen Operator mit eingliedriger Koeffizientenfolge zu reduzieren. Die verbleibende Ergänzung einer semi–unitären Matrix zu einer unitären Matrix kann mit Standardmethoden der linearen Algebra vorgenommen werden.

Satz 3.6.1 (vgl. [SHGB93], Fact 2) *Seien V, W endlichdimensionale hermitesche Vektorräume. Jede semi–unitäre $(1,1)$–periodische lineare Abbildung $F : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(W)$ kann als Produkt*

$$F = \mathcal{T}^{-M} M (I - P_N(1-\mathcal{T})) \circ \cdots \circ (I - P_1(1-\mathcal{T})) \circ U$$

mit $M, N \in \mathbb{N}$, orthogonalen Projektoren $P_1, \ldots, P_N : W \to W$ und einer semi–unitären Abbildung $U : V \to W$ dargestellt werden.

Beweis: Seien $M, N \in \mathbb{Z}$ und $f_0, \ldots, f_N : V \to W$ so gewählt, dass nach Satz 3.1.7 und mit einer offensichtlichen Umnummerierung die Darstellung

$$F = \mathcal{T}^{-M} f(\mathcal{T}) = f_0 \mathcal{T}^{-M} + \cdots + f_N \mathcal{T}^{N-M}, \qquad f_0 \neq 0,\ f_N \neq 0$$

gilt. Sei weiter für jedes $z \in \mathbb{C}$ die Abbildung $f(z) := f_0 + z f_1 + \cdots + z^N f_N : V \to W$ definiert. Nach Lemma 3.2.4 ist diese Abbildung für alle $z \in S^1$, d.h. für $|z| = 1$, semi–unitär.

Nach demselben Lemma folgt ebenso $f_0^* \circ f_N = 0$, d.h. das Bild von F_0 steht senkrecht zum Bild von F_N. Sei $P_N : W \to W$ der orthogonale Projektor auf dem Bildraum $im F_N$. Dann gelten $P_N^* = P_N$ und $(I - P_N) \circ P_N = 0$. Da nach Konstruktion $P_N \circ F_N = F_N$ und $P_N \circ F_0 = 0$ gelten, ist

$$(I - P_N(1-z^{-1}))f(z) = \big(f_0 + P_N \circ f_1\big) + \big((I-P_N)\circ f_1 + P_N \circ f_2\big)z + \ldots \\ \cdots + \big((I-P_N)\circ f_{N-1} + f_N\big)z^{N-1}$$

ein Ausdruck geringeren Grades in z. Da die Bilder des im konstanten Glied hinzukommenden Ausdrucks $P_N \circ F_1$ senkrecht zu den Bildern von F_0 stehen, ist das konstante Glied immer noch von Null verschieden. Analog dazu ist das Glied höchsten Grades von Null verschieden. Nach einer endlichen Anzahl von Wiederholungen dieses Arguments erhalten wir orthogonale Projektoren $P_1, \ldots, P_N$, mit welchen

$$U := (I - P_1(1-z^{-1})) \circ \cdots \circ (I - P_N(1-z^{-1}))f(z)$$

von z unabhängig ist. Setzen wir in dieser Identität $z := 1$, so ergibt sich die Identität $U = f(1) = f_0 + \cdots + f_N$. Da nun die Faktoren $(I - P_k(1 - z^{-1}))$ unitär sind, gilt

$$f(z) = (I - P_N(1 - z)) \circ \cdots \circ (I - P_1(1 - z)) \circ U \,.$$

Diese polynomiale Identität bleibt erhalten, wenn wir z durch $\mathcal{T}$ ersetzen, woraus die Behauptung folgt. □

Korollar 3.6.2 (vgl. [SHGB93]) *Eine lineare $(1,1)$–periodische Abbildung $F : \ell_{\text{fin}}(\mathbb{C}) \to \ell_{\text{fin}}(\mathbb{C}^p)$ ist genau dann semi–unitär, wenn es normierte Spaltenvektoren $v_0, v_1, \ldots, v_L \in \mathbb{C}^p$, $\|v_k\|_2 = 1$, gibt mit*

$$F = \big(I_p - \mathbf{v}_L \mathbf{v}_L^*(1 - \mathcal{T})\big) \cdot \cdots \cdot \big(I_p - \mathbf{v}_1 \mathbf{v}_1^*(1 - \mathcal{T})\big)\, v_0 \,.$$

Beweis: Ist $\mathbf{v} = (v_1, \ldots, v_p)^t \in \mathbb{C}^p$ ein normierter Spaltenvektor, so ist $\mathbf{v}^* = (\bar{v}_1, \ldots, \bar{v}_p)$ der hermitesch konjugierte Zeilenvektor und $\mathbf{v}\mathbf{v}^*$ ist ein orthogonaler Projektor. Andererseits läßt sich zu jedem orthogonalen Projektor P auf $\mathbb{C}^p$ eine orthogonale Basis $\mathbf{v}_1, \ldots, \mathbf{v}_r$ des Bildraumes des Projektors finden, mit welcher $P = \mathbf{v}_1\mathbf{v}_1^* + \cdots + \mathbf{v}_r\mathbf{v}_r^*$ gilt. Dann ist jedoch ebenso

$$I_p - P(1 - z) = \big(I_p - \mathbf{v}_r \mathbf{v}_r^*(1 - \mathcal{T})\big) \cdot \cdots \cdot \big(I_p - \mathbf{v}_1 \mathbf{v}_1^*(1 - \mathcal{T})\big) \,,$$

jeder der Faktoren, welche sich aus der Zerlegung nach dem vorhergehenden Satz ergeben, kann also weiter in Faktoren zu Rang–1–Projektoren zerlegt werden. □

3.6.2 Faktorisierung symmetrischer semi–unitärer Operatoren

Lemma 3.6.3 *Seien V ein euklidischer Vektorraum mit Symmetrie $J : V \to V$ und $a \in V_+$, $b \in V_-$ je ein gerader und ungerader Einheitsvektor. Mit*

$$P := \frac{1}{2}(a + b)(a^* + b^*) \text{ und } R = R(a,b) := I + P(\mathcal{T} - 1) + JPJ(\mathcal{T}^{-1} - 1) \,.$$

ist P ein orthogonaler Projektor auf V, $R : \ell_{\text{fin}}(V) \to \ell_{\text{fin}}(V)$ ist unitär und es gilt $\tau_\nu R \tau_\nu = JRJ = R(a, -b) = R^{-1}$ für jedes $\nu \in \mathbb{Z}$.

Beweis: Da a und b orthogonal zueinander sind, gilt $\langle\, a + b,\, a + b \,\rangle = 2$. Somit gilt $P^2 = P$, nach Konstruktion gilt $P^* = P$. Weiter ist $\tau_\nu \mathcal{T} \tau_\nu = \mathcal{T}^{-1}$, woraus aus der Struktur von R die Identität $\tau_\nu R \tau_\nu = JR(a,b)J$ folgt. Um die Orthogonalität nachzuweisen, sei als erstes bemerkt, dass mit $J(a + b) = a - b$ auch $JPJ = \frac{1}{2}(a - b)(a^* - b^*)$ ist. Wegen $\langle\, a + b,\, a - b \,\rangle_V = 0$ gilt nun $JPJP = 0$, somit ist

$$(I - P(\mathcal{T} - 1))(I - JPJ(\mathcal{T}^{-1} - 1)) = R \,.$$

Jeder Faktor für sich ist jedoch unitär, somit auch die Verknüpfung. Da die Projektoren selbstadjungiert sind, ist $R(a,b)^{-1} = R^* = \tau_\nu R \tau_\nu = JRJ = R(a,-b)$. □

Satz 3.6.4 (vgl. [Tur94]) *Seien V ein endlichdimensionaler euklidischer Vektorraum und $J : V \to V$ eine lineare Abbildung mit $J^2 = 1$. Sei $F : \ell_{\text{fin}}(\mathbb{R}) \to \ell_{\text{fin}}(V)$ eine lineare Abbildung, die*

a) $(1,1)$–periodisch,

b) semi–unitär und

c) symmetrisch in dem Sinne ist, dass für $\nu = 0$ oder $\nu = 1$ die Identität $\tau_\nu \circ F \circ \tau_\nu = JF$ gilt.

Dann gibt es ein $M \in \mathbb{N}$ und gerade Einheitsvektoren $a_1, \ldots, a_M, c \in V_+$ sowie ungerade Einheitsvektoren $b_1, \ldots, b_M, d \in V_-$, so dass für $\nu = 0$

$$F = R(a_M, b_M) \cdot \cdots \cdot R(a_1, b_1) \cdot c$$

und für $\nu = 1$

$$F = R(a_M, b_M) \cdot \cdots \cdot R(a_1, b_1) \cdot \left(\tfrac{1}{2}(1+\mathcal{T})c + \tfrac{1}{2}(1-\mathcal{T})d\right)$$

gilt.

Beweis: Seien $M, N \in \mathbb{N}$ und $f_{-M}, \ldots, f_{N-M} \in V$, $f_{N-M} \neq 0$ die Vektoren, mit welchen $F = f_{-M}\mathcal{T}^{-M} + \cdots + f_{N-M}\mathcal{T}^{N-M}$ gilt. Aus der Symmetrie laut Voraussetzung c) ergibt sich, dass dann $f_k = Jf_{\nu-k}$ gilt, insbesondere muss für die Indexgrenzen $N = \nu + 2M$ gelten. Da J unitär auf V ist, gilt ebenso $\|f_k\|_V = \|f_{\nu-k}\|_V$.

Die nur die jeweils äußeren beiden Glieder betreffenden Bedingungen aus der Semi–Unitarität von F lauten

$$\begin{aligned} 0 &= \langle\, f_{-M},\, f_{\nu+M} \,\rangle_V = \langle\, Jf_{\nu+M},\, f_{\nu+M} \,\rangle_V \text{ und} \\ 0 &= \langle\, f_{-M},\, f_{\nu+M-1} \,\rangle_V + \langle\, f_{-M+1},\, f_{\nu+M} \,\rangle_V \\ &= \langle\, Jf_{\nu+M},\, f_{\nu+M-1} \,\rangle_V + \langle\, Jf_{\nu+M-1},\, f_{\nu+M} \,\rangle_V \ . \end{aligned}$$

Da der Vektorraum V reell ist, ist das Skalarprodukt symmetrisch, somit steht $f_{\nu+M}$ sowohl zu $Jf_{\nu+M}$ als auch zu $Jf_{\nu+M-1}$ senkrecht. Seien $r, s > 0$ und $a \in V_+$, $b \in V_+$ Einheitsvektoren mit $f_{\nu+M} = ra + sb$. Da $Jf_{\nu+M} = ra - sb$ gilt, folgt neben $\|f_{\nu+M}\|_V^2 = r^2 + s^2$ auch $\|ra\|_V = \|sb\|_V$, d.h. $r = s$.

Mit $P = \frac{1}{2}(a+b)(a^* + b^*)$ gelten also

$$\begin{aligned} P(f_{\nu+M}) &= \tfrac{r}{2}(a+b)(a^*+b^*)(a+b) = f_{\nu+M} \\ JPJ(f_{\nu+M}) &= \tfrac{r}{2}(a-b)(a^*-b^*)(a+b) = 0 \\ JPJ(f_{\nu+M-1}) &= \tfrac{1}{2r}(a-b)\,\langle\, J(f_{\nu+M}),\, f_{\nu+M-1} \,\rangle_V = 0\,. \end{aligned}$$

Die Verknüpfung von $R(a,-b) = I + JPJ(\mathcal{T}-1) + P(\mathcal{T}^{-1}-1)$ mit F ergibt somit einen semi–unitären, symmetrischen Differenzenoperator

$$\tilde{F} := R(a,-b) \circ F = \tilde{f}_{-M+1}\mathcal{T}^{-M+1} + \cdots + \tilde{f}_{d\nu+M-1}\mathcal{T}^{\nu+M-1}$$
$$\text{mit } \tilde{f}_k = Pf_{k+1} + (I - P - JPJ)f_k + JPJf_{k-1}\,,$$

welcher in beiden Richtungen einen Summanden weniger aufweist. Wir setzen also $a_M := a$ und $b_M = b$ und wiederholen diesen Schritt, bis Einheitsvektoren $a_1, \ldots, a_M \in V_+$ und $b_1, \ldots, b_M \in V_-$ gefunden sind, mit welchen

$$R(a_1, -b_1) \cdot \ldots \cdot R(a_M, -b_M)\, F = \begin{cases} g_0 & \text{bei } \nu = 0 \\ g_0 + g_1 \mathcal{T} & \text{bei } \nu = 1 \end{cases}$$

gilt. Bei $\nu = 0$ muss $g_0 = Jg_0$ gelten. Da die Verknüpfung weiter semi–unitär ist, gilt auch $\|g_0\|_V = 1$; $c := g_0$ ist also der noch fehlende gerade Einheitsvektor. Bei $\nu = 1$ ist $g_1 = Jg_0$, analog zur Argumentation oben sind $c := g_0 + g_1 \in V_+$ und $d := g_0 - g_1 \in V_-$ Einheitsvektoren, es gilt $g_0 + g_1\mathcal{T} = \frac{1}{2}(1 + \mathcal{T})c + \frac{1}{2}(1 - \mathcal{T})d$. Da $R(a, -b)$ invers zu $R(a, b)$ ist, folgt die Behauptung des Satzes. □

Kapitel 4

Abtastung und Interpolation

In vielen Problemen der Natur– und Ingenieurwissenschaften wird ein zeitkontinuierlicher physikalischer Prozess mittels in der Zeit stetiger Funktionen modelliert. Die Räume derartiger Funktionen sind aber stets unendlich-dimensional.

Um Berechnungen mit dem Modell – welcher Art auch immer – durchführen zu können, bedarf es endlich–dimensionaler Räume, d.h. das berechenbare Modell darf nur von endlich vielen Parametern abhängen.

Will man die Tauglichkeit eines Modells am real ablaufenden Prozess überprüfen, braucht man eine Methode, die es gestattet, die Parameter zu bestimmen. Dazu sind für den Prozess typische Größen zu messen. Dies kann auch nur in endlicher Weise geschehen, d.h. es kommen nur endlich viele Werte mit endlicher Präzision in Frage.

Um den Verlauf des Prozesses in der Zeit zu erfassen, wiederholt man die Messung der Größen mehrfach zu aufeinander folgenden Zeitpunkten. Auf diese Weise entsteht eine Messreihe, *Zeitreihe* genannt.

Die Aufstellung einer Zeitreihe nennt man in der Signaltheorie *Abtastung* des Prozesses.

Im Allgemeinen ist der Messwert zu einem gewissen Zeitpunkt eine Funktion des Zustandes eines vorliegenden Prozesses in diesem Zeitpunkt. Lassen wir das Problem der Rekonstruktion des Zustandes aus den Messwerten außer acht und nehmen vereinfachend an, dass ein einfacher Messwert schon den Zustand charakterisiert. Nehmen wir weiter idealisierend an, dass fehlerfrei gemessen werden kann. So bleibt immer noch das Problem, den endlich vielen Paaren (bestehend jeweils aus Zeitpunkt und Messwert), *Stützstellen* genannt, eine stetige Funktion zuzuordnen.

Zu dessen Lösung wurden verschiedenste Interpolationsverfahren entwickelt. Die ersten und einfachsten Verfahren nach Newton und Lagrange bestimmen ein Polynom genügend hohen Grades, welches durch die Stützstellen verläuft. Sollen jedoch neue Messwerte hinzugefügt werden, so muss die interpolierende Funktion völlig neu bestimmt werden. Außerdem kann sich die Interpolationsfunktion bei solch einer Änderung sehr stark ändern.

Abhilfe für Messzeitpunkte gleichen Abstands schafft die Interpolation mit dem Kardinalsinus. Die Betrachtung der mit diesem konstruierten Kardinalreihen führt auf die Untersuchung von Fourier–Reihen, der Fourier–Transformation und die Zerlegung der Zeit–Frequenz–Ebene auf verschiedene Arten.

Mit Hilfe dieser Werkzeuge wurde die moderne digitale Nachrichtentechnologie geschaffen ([Sha49]). Das Grundprinzip der digitalen Kommunikation läßt sich folgendermaßen skizzieren: Eine endliche Folge von Zahlen (meist aus einem endlichen Wertevorrat) wird einem ebensolangen Abschnitt einer Folge äquidistanter Zeitpunkte als Werte zugeordnet. Der zeitliche Abstand in der Folge wird auch als Taktlänge bezeichnet. Zu diesen Paaren von Zeitpunkt und Wert wird eine Interpolationsfunktion konstruiert, die als Spannung eines elektrischen Leiters oder drahtlos als Amplitude eines hochfrequenten elektromagnetischen Feldes realisiert wird. Am anderen Ende des Leiters oder einem anderen Ort in Reichweite des elektromagnetischen Wechselfeldes wird diese Funktion dann in Form einer Folge von Messwerten gemessen, abgetastet.

Dabei können in der Zeit–Frequenz–Ebene Frequenzbänder festgelegt werden, die als unabhängige Kommunikationskanäle dienen. Denn sind (bis auf Randpunkte) disjunkte Intervalle auf der Frequenzachse festgelegt, so kann durch Modulation der Kardinalreihe eine Interpolationsfunktion geschaffen werden, deren Fourier–Transformierte auf dieses Intervall beschränkt bleibt, d.h. deren Darstellung in der Zeit–Frequenz–Ebene nur den diesem Intervall entsprechenden Streifen belegt. Es gibt für jedes Intervall eine von dessen Breite abhängige minimale Taktlänge. Durch diese wird die in Frequenzbänder zerlegte Ebene auch in der Zeit unterteilt.

Es wird sich herausstellen, dass man zur idealen Umsetzung dieser Theorie für jeden Messwert Integrale bestimmen oder approximieren muss, die sich über die gesamte Zeitachse erstrecken. Aus Methoden, diesen Nachteil bzw. den Fehler bei der Approximation systematisch zu fassen und daraus den Vorteil schneller Algorithmen zu gewinnen, entstanden verschiedene Verfahren der gefensterten Fourier–Transformation und schließlich die Wavelet–Theorie mit ihren Methoden insbesondere der Multi–Skalen–Analyse (s. [Mal89, Mey89, Dau96]).

4.1 Interpolation

Die Problemstellung, auf welche Interpolationsverfahren eine Antwort geben, ist die folgende: Seien $N \in \mathbb{N}$ Paare $(x_n, y_n) \in \mathbb{K}^2$, $n = 1 \ldots, N$ gegeben, wobei $\mathbb{K} = \mathbb{R}$ der Körper der reellen Zahlen oder $\mathbb{K} = \mathbb{C}$ der Körper der komplexen Zahlen sein kann sowie $x_1, \ldots, x_N$ paarweise verschieden sind. Zu diesen soll eine Funktion $F : \mathbb{K} \to \mathbb{K}$ gefunden werden, welche $F(x_n) = y_n$ für jedes $n = 1 \ldots, N$ erfüllt.

4.1.1 Interpolationskerne aus Differenzenquotienten

Die Lösung dieser Aufgabe erweist sich als besonders einfach, wenn Funktionen $L_k : \mathbb{K} \to \mathbb{K}$ bekannt sind, für welche $L_k(x_n) = \delta_{k,n}$ mit dem Kronecker–Symbol $\delta_{k,n}$ gilt, d.h. an allen Punkten x_n, $n = 1, \ldots, N$ hat L_k den Wert Null, ausgenommen den Punkt x_k, an welchem

der Wert 1 angenommen wird. Solche Funktionen werden *Interpolationskernfunktionen* oder *Interpolationskerne* genannt.

Eine Möglichkeit, sich solche Interpolationskerne L_k zu verschaffen, besteht darin, von einer stetig differenzierbaren (bzw. holomorphen) Funktion $H : \mathbb{K} \to \mathbb{K}$ auszugehen, welche unter ihren einfachen Nullstellen die Punkte $x_1, \ldots, x_N$ aufweist. Dabei ist eine Nullstelle $z \in \mathbb{K}$, $H(z) = 0$ einfach, falls die Ableitung dort nicht verschwindet, $H'(z) \neq 0$. Dann ist ein Interpolationskern definiert als Quotient der Funktion H und ihrer Linearisierung in x_k:

$$L_k(x) := \begin{cases} 1 & x = x_k, \\ \frac{H(x)}{H'(x_k)(x-x_k)} & x \neq x_k \end{cases} . \tag{4.1}$$

Die Nullstellen außer x_k bleiben dabei erhalten, in x_k hat diese Funktion nach Definition den Wert 1 und ist nach dem Mittelwertsatz im Punkte x_k auch stetig. Somit erhalten wir als interpolierende Funktion für die Paare (x_k, y_k), $k = 1, \ldots, N$

$$F := \sum_{k=1}^{N} y_k \, L_k \, .$$

Um nun eine Funktion H mit paarweise verschiedenen einfachen Nullstellen $x_1, \ldots, x_N$ zu konstruieren, kann man von einer Funktion $h : \mathbb{K} \to \mathbb{K}$ mit einer einfachen Nullstelle im Nullpunkt ausgehen, $h(0) = 0$ und $h'(0) \neq 0$. Mit dieser wird das Produkt

$$H(x) := \prod_{n=1}^{N} h(x - x_n)$$

gebildet, welches somit Nullstellen in den Punkten $x_1, \ldots, x_N$ besitzt. Um die Einfachheit der Nullstellen zu sichern, darf h nicht auf den von Null verschiedenen Differenzen der Stützpunkte verschwinden, $h(x_k - x_n) \neq 0$ für alle $k, n = 1 \ldots, N$ mit $k \neq n$. Dann erhalten wir für die Ableitung in den Nullstellen den Ausdruck

$$\begin{aligned} H'(x_k) = &\sum_{m=1}^{N} h'(x_k - x_m) \prod_{n \neq m} h(x_k - x_n) \\ = &h'(0) \prod_{n \neq k} h(x_k - x_n) \end{aligned}$$

und somit für die Interpolationskerne

$$L_k(x) := \frac{h(x - x_k)}{h'(0)(x - x_k)} \prod_{n \neq m} \frac{h(x - x_n)}{h(x_k - x_n)} \, . \tag{4.2}$$

4.1.2 Beispiele für endliche Interpolationsmethoden

Beispiele für Interpolationsverfahren, die sich nach diesem Schema konstruieren lassen, sind:

1. Die Lagrange–Interpolation, welche sich als einfachste Variante mit $h(x) := x$ ergibt und auf eine interpolierende Funktion der folgenden Art führt:

$$F(x) := \sum_{k=0}^{N} y_k \prod_{n \neq k} \frac{x - x_n}{x_k - x_n} \, .$$

2. Die Wahl einer nichtlinearen Funktion wie beispielsweise $h(x) := x/(1 + W^2 x^2)$ mit $h'(0) = 1$ ergibt auf $\mathbb{K} = \mathbb{R}$ eine „gedämpfte“ Interpolation

$$F(x) := \sum_{k=1}^{N} \frac{y_k}{1 + W^2(x - x_k)^2} \prod_{n \neq k} \frac{x - x_n}{x_k - x_n} \frac{1 + W^2(x_k - x_n)^2}{1 + W^2(x - x_n)^2} \, .$$

3. Die Interpolation mittels Winkelfunktionen, die mit der *diskreten Fourier–Transformation* (DFT) verbunden ist, ergibt sich als Spezialfall der Lagrange–Interpolation zu den Einheitswurzeln eines Grades N. Diese sind die Nullstellen des Kreisteilungspolynoms $H(x) := x^N - 1$ und seien durch $x_k = \exp(i\frac{2\pi}{N}k)$, $k = 1, \ldots, N$, aufgezählt. Nach Polynomdivision ergibt sich

$$L_k(x) := \frac{x^N - x_k^N}{N x_k^{N-1}(x - x_k)} = \frac{(\bar{x}_k x)^N - 1}{N(\bar{x}_k x - 1)} = \frac{1}{N}(1 + \bar{x}_k x + \cdots + (\bar{x}_k x)^{N-1})$$

und somit durch Vertauschen der Summen in der interpolierenden Funktion

$$F(x) := \sum_{k=1}^{N} y_k L_k(x) = \sum_{n=0}^{N-1} x^n \cdot \frac{1}{N} \sum_{k=1}^{N} y_k \bar{x}_k^n \, .$$

Die darin auftretenden Koeffizienten

$$\hat{y}_n := \frac{1}{N} \sum_{k=1}^{N} y_k \exp(-i\tfrac{2\pi}{N}kn) \tag{4.3a}$$

entsprechen dem Ergebnis der *diskreten Fourier–Transformation* (DFT). Die Auswertung des Interpolationspolynoms an den Nullstellen entspricht der inversen DFT

$$y_k = F(x_k) = \sum_{n=0}^{N-1} e^{i\frac{2\pi}{N}kn} \hat{y}_n \, . \tag{4.3b}$$

4. Die Interpolation zu ganzzahligen Stützstellen ergibt sich als Grenzfall der Lagrange–Interpolation. Seien dazu zunächst die Stützstellen ganzzahlig im Intervall $\{-N, \ldots, N\}$ gewählt. Nach dem zur Lagrange–Funktion gesagten ist eine Hilfsfunktion, welche gerade diese Nullstellen hat, die Funktion

$$H_N(x) := \frac{(-1)^N}{(N!)^2}(x + N) \cdots (x - N) = x \prod_{n=1}^{N} (1 - \frac{x^2}{N^2}) \, .$$

Als weitere Funktion, welche die ganzen Zahlen als Nullstellen hat, bietet sich der – um einen Faktor π skalierte – Sinus an. Dieser ergibt sich sogar als Grenzwert für unendlich großes N, genauer gesagt gilt $\lim_{N\to\infty} H_N(x) = \sin(\pi x)/\pi$. Mit $H(x) := \sin(\pi x)$

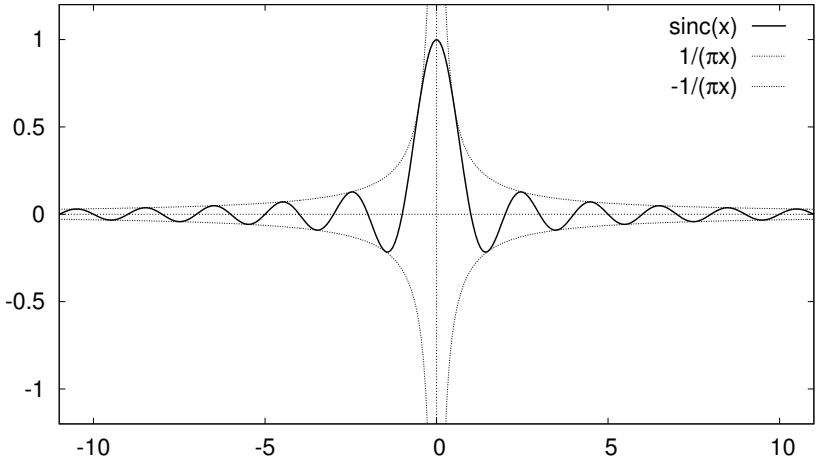


Abbildung 4.1: Der *Kardinalsinus* mit seiner hyperbolischen Hülle

erhalten wir die (1915 von E. T. Whittaker [Whi15] so benannte) *Kardinalreihe*

$$F(x) = \sum_{k\in I} y_k \frac{\sin(\pi x)}{\pi\cos(\pi k)(x-k)} = \frac{\sin \pi x}{\pi} \sum_{k\in I} \frac{y_k}{(-1)^k (x-k)}$$
$$= \sum_{k\in I} y_k \frac{\sin(\pi(x-k))}{\pi(x-k)} \,.$$

Dabei ist hier $I \subset \mathbb{Z}$ eine endliche Indexmenge. Der Interpolationskern, welcher hier in ganzzahligen Verschiebungen auftritt, wird *Sinus cardinalis* bzw. *Kardinalsinus* genannt und mit $\operatorname{sinc}(x) := \frac{\sin \pi x}{\pi x}$ bezeichnet. Diese Funktion hat verschiedene nutzvolle Darstellungen, darunter die aus dem Sinus abgeleitete Potenzreihe

$$\operatorname{sinc}(x) = 1 + \sum_{n=1}^{\infty} \frac{(-\pi^2 x^2)^k}{(2k+1)!} \,,$$

welche auf ganz $\mathbb{R}$ konvergiert und aus der man auch die symmetrische Natur dieser Funktion ersehen kann. Desweiteren gibt es eine im weiteren wichtige Integraldarstellung

$$\operatorname{sinc}(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi s x} \,\mathrm{d}s = 2\int_{0}^{\frac{1}{2}} \cos(2\pi s x) \,\mathrm{d}s \,.$$

Auf die Konvergenz der Kardinalreihe bei unendlich vielen Stützstellen wird im nächsten Abschnitt eingegangen.

5. Sei $h(x) := \sin(\Omega x)$. Sind $a_1, \ldots, a_N$ so gewählt, dass unter den paarweisen Differenzen $a_k - a_m$ sich keine ganzzahligen Vielfachen von $T := \frac{\pi}{\Omega}$ befinden, so hat die Funktion

$$H(x) := h(x-a_1)\cdots h(x-a_N)$$

die einfachen Nullstellen $a_k + nT$, $k = 1, \ldots, N$, $n \in \mathbb{Z}$. Ist $I \subset \{1, \ldots, N\} \times \mathbb{Z}$ eine endliche Teilmenge, und zu jedem Indexpaar $(k, n) \in I$ ein Wert $y_{n,k} \in \mathbb{K}$ vorgegeben,

1914–15.] **Expansions of the Interpolation-Theory.** 187

or

$$\frac{w}{\pi}\sin\frac{\pi}{w}(x-a)\sum_{r=-\infty}^{\infty}\frac{(-1)^r f(a+rw)}{x-a-rw}. \quad . \quad . \quad . \quad (4)$$

represents a function which is cotabular with the given function f(x), *but which has no periodic constituents of period less than* 2w.

Now, in order to construct the expression (3) or (4), we do not need to know anything about $f(x)$ except its values $f(a)$, $f(a+w)$, $f(a-w)$, etc., at the tabulated values of the argument. These values, however, are not peculiar to $f(x)$, but are common to the whole set of cotabular functions. It follows that *we arrive at the same expression* (3) *whatever function* f(x) *of the cotabular set we start from.* The expression (3) is therefore an *invariantive function* of the cotabular set: and it may be regarded as the *simplest function* belonging to the set. We shall call it the CARDINAL FUNCTION of the set.

Abbildung 4.2: Ausschnitt aus [Whi15] von Edmund Taylor Whittaker

so ergibt sich die Interpolationsfunktion

$$F(x) := \sum_{(k,n)\in I} y_{n,k}\frac{\sin(\Omega(x-nT-a_k))}{\Omega(x-nT-a_k)}\prod_{m\neq k}\frac{\sin(\Omega(x-nT-a_m))}{\sin(\Omega(a_k-a_m))}\,.$$

4.1.3 Die Kardinalreihe

Edmund T. Whittaker [Whi15] (nach [Hig85]) betrachtete das Problem, zu einer gegebenen komplexwertigen Folge $\{a_n\}_{n\in\mathbb{Z}}$ holomorphe Funktionen $f : \mathbb{C} \to \mathbb{C}$ mit $f(nT) = a_n$ für jedes $n \in \mathbb{Z}$ zu finden. Die Menge aller dieser Funktionen nannte er *kotabulare Menge*, in dieser zeichnet sich die auf unendliche Folgen verallgemeinerte Kardinalreihe, sofern sie konvergiert, durch die geringste „Schwankung“ aus.

Die Kardinalreihe an sich war schon etwas früher bekannt, so findet sich bei Emile Borel ([Bor99] nach [Hig85]) folgende Konvergenzbedingung:

Lemma 4.1.1 *Ist eine Zahlenfolge $\{a_n\}_{n\in\mathbb{Z}}$ gegeben, so konvergiert die Kardinalreihe*

$$F(x) := \sum_{n\in\mathbb{Z}} a_n \operatorname{sinc}(x-n)$$

genau dann punktweise und lokal gleichmäßig auf ganz $\mathbb{R}$, *wenn*

$$\sum_{n\in\mathbb{Z}\setminus\{0\}}\left|\frac{a_n}{n}\right| < \infty\,.$$

Beweis: Seien ein $N \in \mathbb{N}$ und ein $x \in (-N, N)$ fixiert. Dann gilt für alle $n \in \mathbb{Z}$ mit $|n| > 2N$

$$2n \geq 2N + n \geq |x-n| \geq |n| - |x| \geq |n| - N \geq |n|/2\,.$$

Somit gilt auch

$$\frac{1}{2}\left|\frac{a_n}{n}\right| \le \left|\frac{a_n}{x-n}\right| \le 2\left|\frac{a_n}{n}\right| ,$$

und wir erhalten die Abschätzung der Kardinalreihe

$$\sum_{n=-2N}^{2N} |a_n \operatorname{sinc}(x-n)| + \frac{|\sin \pi x|}{2\pi} \sum_{|n|>2N} \left|\frac{a_n}{n}\right| \le \sum_{n\in\mathbb{Z}} |a_n \operatorname{sinc}(x-n)| \le \sum_{n=-2N}^{2N} |a_n \operatorname{sinc}(x-n)| + \frac{2}{\pi} \sum_{|n|>2N} \left|\frac{a_n}{n}\right| .$$

Diese beidseitige Abschätzung besagt, dass die Reihe $\sum_{|n|>1} \left|\frac{a_n}{n}\right|$ sowohl Majorante als auch Minorante für die absolute Konvergenz der Kardinalreihe ist. Somit stimmen deren Konvergenz– bzw. Divergenzverhalten überein. Konvergiert die Reihe $\sum_{|n|>1} \left|\frac{a_n}{n}\right|$, so konvergiert die Kardinalreihe überall und lokal gleichmäßig, divergiert sie, so ist nirgendwo lokal gleichmäßige Konvergenz gegeben. □

Die Voraussetzung dieses Satzes sind für alle Folgen aus den Folgenräumen $\ell_1(\mathbb{C})$ der absolut summierbaren Folgen und $\ell_2(\mathbb{C})$ der im Betragsquadrat summierbaren Folgen erfüllt (zur Definition dieser Räume s. Anhang A.1). Für $a \in \ell_1(\mathbb{C})$ ist dies sofort einsichtig, für $a \in \ell_2(\mathbb{C})$ folgt die Gültigkeit der Voraussetzung aus der Endlichkeit der Norm $\|a\|_2^2 := \sum_{n\in\mathbb{Z}} |a_n|^2 < \infty$ und der Cauchy–Schwarzschen Ungleichung

$$\sum_{|n|>0} \left|\frac{a_n}{n}\right| \le \sqrt{2\sum_{n=1}^{\infty} \frac{1}{n^2}} \sqrt{\sum_{|n|>0} |a_n|^2} \le \frac{\pi}{\sqrt{3}} \|a\|_2 < \infty .$$

Die Integraldarstellung des Kardinalsinus kann in die Kardinalreihe eingesetzt werden,

$$F(x) = \sum_{n\in\mathbb{Z}} a_n \operatorname{sinc}(x-n) = \sum_{n\in\mathbb{Z}} \int_{-\frac{1}{2}}^{\frac{1}{2}} a_n e^{i\,2\pi\, s(x-n)} \,\mathrm{d}s .$$

Ist die Reihe zur Folge $a = \{a_n\}$ absolut konvergent, $a \in \ell_1(\mathbb{C})$, so kann in obiger Formel das Integral mit der Reihe vertauscht werden. Der Integrand entspricht dann einer *trigonometrischen Fourier–Reihe* (s. Anhang B)

$$\mathcal{E}(a)(s) := \sum_{n\in\mathbb{Z}} a_n e^{i2\pi\, sn} .$$

Die Fourier–Reihe konvergiert punktweise absolut und als Funktionenreihe gleichmäßig, die Funktion im Grenzwert ist also stetig.

Somit hat die Kardinalreihe zu $a \in \ell_1(\mathbb{C})$ auch die Darstellung

$$F(x) = \int_I \sum_{n\in\mathbb{Z}} a_n e^{i2\pi\,\omega(x-n)} \,\mathrm{d}\omega = \int_I g(\omega) e^{i2\pi\,\omega x} \,\mathrm{d}\omega = \int_{\mathbb{R}} \chi_I(\omega) g(\omega) e^{i2\pi\,\omega x} \,\mathrm{d}\omega ,$$

wobei $g(\omega) = \mathcal{E}(a)(-\omega)$ gesetzt wurde und χ_I die Indikatorfunktion des Intervalls $I = \left[-\frac{1}{2}, \frac{1}{2}\right]$ ist, d.h. $\chi_I(\omega)$ hat den Wert 1 für $\omega \in I$, den Wert 0 im gegenteiligen Fall.

Diese letzte Darstellung von F entspricht der inversen Fourier–Transformation, $F = \mathcal{F}^{-1}(\chi_I g)$. Dabei verwenden wir die Fourier–Transformation nach der „echten“ Frequenz (s. Anhang B.4). Für beliebige $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ ist die Fourier–transformierte Funktion $\hat{f} := \mathcal{F}(f) \in L^2(\mathbb{R})$ definiert als

$$\hat{f}(\omega) = \mathcal{F}(f)(\omega) := \int_{\mathbb{R}} f(x) e^{-i(2\pi\omega)x} \, \mathrm{d}x \, .$$

Die inverse Transformation ist gleichzeitig die adjungierte, es gelten die Identitäten $\mathcal{F}^{-1}(f) = \mathcal{F}^*(f) = \overline{\mathcal{F}(\bar{f})}$ für jedes $f \in L^2(\mathbb{R})$.

Es seien wieder die „Elementarschwingungen“ $e_\alpha : \mathbb{R} \to \mathbb{C}$ als $x \mapsto e_\alpha(x) := e^{i(2\pi\alpha)x}$ definiert. Nach der Theorie der Fourier–Reihen ist das System der Funktionen $\{\chi_I e_n\}_{n\in\mathbb{Z}}$ eine Hilbert–Basis des Funktionenraums $L^2(I) := L^2(I, \mathbb{C})$. Dies bedeutet, dass für jede Folge $a \in \ell_2(\mathbb{C})$ die Funktionenreihe

$$\mathcal{E}(a) := \sum_{n\in\mathbb{Z}} a_n \chi_I e_n$$

in $L^2(I)$ und auch in $L^2(\mathbb{R})$ konvergiert. Weiter kann jede Funktion $g \in L^2(I)$ bzw. $g \in L^2(\mathbb{R})$ mit Träger in I in eine Fourier–Reihe entwickelt werden, d.h. es gibt eine Koeffizientenfolge $a \in \ell_2(\mathbb{C})$ mit $g = \chi_I \mathcal{E}(a)$. Die Gleichheit der Funktion zu ihrer Fourier–Reihe muss im Sinne des Raumes $L^2(I)$ verstanden werden, d.h. als punktweise Gleichheit außerhalb einer Menge vom Maß Null.

Die Koeffizientenfolge ergibt sich gliedweise für jedes $n \in \mathbb{Z}$ durch die Skalarprodukte

$$a_n = \langle\, g,\, e_n \,\rangle_{L^2(I)} = \int_{\mathbb{R}} g(\omega) \chi_I(\omega) e_{-n}(\omega) \, \mathrm{d}\omega = \mathcal{F}^{-1}(\chi_I g)(-n) \, .$$

Definiert man, von $g \in L^2(I)$ ausgehend, die Funktion $F := \mathcal{F}^{-1}(g) \in L^2(\mathbb{R})$, so ist die Folge $\{F(-n)\}_{n\in\mathbb{Z}}$ von Funktionswerten gleichzeitig die Folge der Fourier–Koeffizienten von g. Somit ist $\{f(-n)\}_{n\in\mathbb{Z}} \in \ell_2(\mathbb{C})$ und f ist als Kardinalreihe darstellbar,

$$f(x) = \mathcal{F}^{-1}\left(\chi_I \sum_{n\in\mathbb{Z}} f(n) e_{-n} \right)(x) = \sum_{n\in\mathbb{Z}} f(n) \mathcal{F}^{-1}(\chi_I e_{-n})(x) = \sum_{n\in\mathbb{Z}} f(n) \operatorname{sinc}(x-n) \, .$$

Insbesondere ist f nach Lemma 4.1.1 stetig.

Wir erhalten somit drei bijektive isometrische Zuordnungen,

- von $\ell_2(\mathbb{C})$ zu $L^2(I)$ durch die Fourier–Reihen,
- von $\ell_2(\mathbb{C})$ zu den Kardinalreihen, die auch $L^2(\mathbb{R})$ angehören, und
- von $L^2(I)$ zu diesen Kardinalreihen durch die inverse Fourier–Transformation.

Dabei entspricht die Kombination der ersten und der dritten Zuordnung gerade der zweiten Zuordnung.

Diese Dreiecksbeziehung läßt sich teilweise oder vollständig für beliebige beschränkte und abgeschlossene Teilmengen von $\mathbb{R}$ anstelle von I formulieren. Interessant sind dabei vor allem Teilmengen, die aus endlich vielen paarweise disjunkten, beschränkten und abgeschlossenen Intervallen zusammengesetzt sind.

4.1.4 Das Abtasttheorem

Definition 4.1.2 *Sei $A \subset \mathbb{R}$ eine abgeschlossene beschränkte Teilmenge. Mit $PW(A)$ sei (in Anlehnung an die* Paley–Wiener*-Theorie, vgl. z.B. [Hig85]) der Unterraum von $L^2(\mathbb{R})$ bezeichnet, welcher aus denjenigen Funktionen $f \in L^2(\mathbb{R})$ besteht, für die der Träger* supp $\hat{f}$ *der Fourier–Transformierten $\hat{f} = \mathcal{F}(f)$ in der Menge A enthalten ist. D.h. mit der Indikatorfunktion χ_A der Menge A die Identität $\hat{f} = \chi_A \hat{f}$ besteht. A wird auch als* Frequenzband *von f bezeichnet, f selbst als* bandbeschränkte Funktion*.*

Im Verlaufe der weiteren Konstruktion wird $\hat{f}$ periodisch fortgesetzt und zu dieser Fortsetzung eine trigonometrische Fourier–Reihe bestimmt. Damit dies möglich ist, muss die Periode der Fortsetzung mit dem Frequenzband A verträglich sein.

Definition 4.1.3 *Sei $A \subset \mathbb{R}$ eine abgeschlossene beschränkte Teilmenge. Jedes $B > 0$ heißt* Bandbreite *von A, wenn die paarweisen Durchschnitte der Mengen $nB + A := \{nB + x : \ x \in A\}$, $n \in \mathbb{Z}$, Mengen vom Maß Null sind.*

In allen weiteren Beispielen ist das Frequenzband A eine endliche Vereinigung beschränkter paarweise disjunkter Intervalle. Die Bedingung an den Durchschnitt verschiedener verschobener Kopien von A reduziert sich darauf, dass nur Randpunkte der Intervalle in den Schnittmengen enthalten sein dürfen. Da A als beschränkt vorausgesetzt ist, kann der Durchmesser $(\sup A - \inf A)$ bestimmt werden. Dieser ist immer eine Bandbreite.

Satz 4.1.4 (Abtasttheorem) *Seien $A \subset \mathbb{R}$ beschränkt und abgeschlossen sowie $B > 0$ eine Bandbreite von A. Seien weiter χ_A die Indikatorfunktion von A und $\varphi := \frac{1}{B}\mathcal{F}^{-1}(\chi_A)$.*

Dann ist jede bandbeschränkte Funktion $f \in PW(A)$ stetig und

a) ist bereits durch ihre Abtastfolge $\{f(\frac{n}{B})\}_{n\in\mathbb{Z}}$ *bestimmt, und*

b) besitzt eine Darstellung $f(x) = \sum_{n\in\mathbb{Z}} f(\frac{n}{B})\varphi(x - \frac{n}{B})$.

Die Funktion φ erzeugt also im Sinne der Aussage b) den Unterraum $PW(A)$.

Beweis: A ist beschränkt, es gibt also ein $\Omega > 0$ mit $A \subset [-\Omega, \Omega]$. Damit gilt auch $PW(A) \subset PW([-\Omega, \Omega]) \subset L^2(\mathbb{R})$. Für jedes $f \in PW(A)$ mit Fourier–Transformierter $\hat{f} := \mathcal{F}(f)$ gilt

$$f(x) = \mathcal{F}^{-1}(\hat{f})(x) = \left\langle \hat{f},\, e_{-x} \right\rangle_{L^2([-\Omega,\Omega])} .$$

Das Skalarprodukt auf $L^2([-\Omega, \Omega])$ ist stetig in beiden Argumenten, und die Abbildung $x \mapsto \chi_{[-\Omega,\Omega]} e_x \in L^2([-\Omega, \Omega])$ ist stetig in $x \in \mathbb{R}$. Somit ist auch f stetig.

Seien $f \in PW(A)$ beliebig und $\hat{f} := \mathcal{F}(f)$ die Fourier–Transformierte. Mit der Voraussetzung der Bandbeschränktheit gilt

$$\hat{f} = \chi_A \hat{f} = \chi_A \sum_{n\in\mathbb{Z}} \mathcal{T}_{nB}\hat{f}.$$

Denn die Punkte, für welche zwei Summanden der Reihe gleichzeitig von Null verschieden sind, formen eine Menge vom Maß Null. Es gibt abzählbar viele Paare von Summanden, somit hat die Reihe fast überall höchstens einen von Null verschiedenen Summanden. Durch den

Faktor χ_A wird die Reihe auf den Summanden mit $n = 0$ reduziert, daher folgt die Gleichheit der Funktionen in $L^2(\mathbb{R})$.

In $L^2([0,B])$ ist das Funktionensystem $\{B^{-1/2}e_{n/B}\}_{n\in\mathbb{Z}}$ eine Hilbert–Basis. Die B–periodische Funktion $g := \sum_{n\in\mathbb{Z}} \mathcal{T}_{nB}\hat{f}$ kann über $[0,B]$ in eine Fourier–Reihe entwickelt werden, die dieser Hilbert–Basis entspricht. Diese Fourier–Reihe ist ebenfalls B–periodisch und es gilt

$$g = \frac{1}{B}\sum_{n\in\mathbb{Z}} \langle\, g,\, e_{n/B}\,\rangle_{L^2([0,B])}\, e_{n/B}\,.$$

Die Fourier–Koeffizienten können weiter umgeformt werden zu

$$\begin{aligned}\langle\, g,\, e_{n/B}\,\rangle_{L^2([0,B])} &= \sum_{k\in\mathbb{Z}} \left\langle\, \mathcal{T}_{kB}\hat{f},\, e_{n/B}\,\right\rangle_{L^2([0,B])} = \sum_{k\in\mathbb{Z}} \left\langle\, \hat{f},\, e_{n/B}\,\right\rangle_{L^2([kB,(k+1)B])} \\ &= \left\langle\, \hat{f},\, e_{n/B}\,\right\rangle_{L^2(\mathbb{R})} = \mathcal{F}(\hat{f})\Big(\frac{n}{B}\Big) = f\Big(-\frac{n}{B}\Big)\end{aligned}$$

Somit gilt für die Fourier–Transformierte

$$\hat{f} = \sum_{n\in\mathbb{Z}} f\Big(\frac{n}{B}\Big)\,\frac{1}{B}\chi_A e_{n/B}\,,$$

sie ist also durch die Folge $\{f(n/B)\}_{n\in\mathbb{Z}}$ eindeutig bestimmt. Nach Anwenden der inversen Fourier–Transformation folgt

$$f = \sum_{n\in\mathbb{Z}} f\Big(\frac{n}{B}\Big)\frac{1}{B}\mathcal{F}^{-1}(\chi_A e_{-n/B}) = \sum_{n\in\mathbb{Z}} f\Big(\frac{n}{B}\Big)\mathcal{T}_{n/B}\varphi\,.$$

□

Das Abtasttheorem kann zu mehreren wichtigen Beispielen spezialisiert werden. Im einfachsten Fall ist die Menge A ein Intervall, spezieller $A = [-\Omega,\Omega]$ mit einem $\Omega > 0$. Es ergibt sich die klassische Formulierung des Abtasttheorems, welches C. E. Shannon (zwischen 1940 und 1949, [Sha49]) nach der Theorie der Kardinalreihen von E. T. Whittaker (1915, [Whi15]) formulierte. [1] Unabhängig davon wurde ein ähnlich lautendes Ergebnis von Kotelnikow (1939) formuliert (nach [Hig85]). Nach den Initialen der Nachnamen dieser drei Forscher, in zeitlicher Reihenfolge gesehen, wird es *WKS–Abtasttheorem* genannt.

Satz 4.1.5 (WKS–Abtasttheorem) *Jede bandbeschränkte Funktion $f \in L^2(\mathbb{R})$ mit maximaler Frequenz $\Omega > 0$ ist*

a) bereits durch die Abtastfolge $\{f(nT)\}_{n\in\mathbb{Z}}$ *mit Schrittweite T bestimmt, falls $T > 0$ auch der Bedingung $2T\Omega \le 1$ genügt, und besitzt*

[1] C. E. Shannon führt diesen Satz in [Sha49] zur Modellierung eines idealen bandbeschränkten Kommunikationskanals ein. Ein Sender erzeugt ein Signal $\sum_{n\in\mathbb{Z}} a_n \operatorname{sinc}(t/T - n)$ aus einer, üblicherweise endlichen, Folge $a = \{a_n\}_{n\in\mathbb{Z}}$ auf eine Weise, dass dieses Signal von einem Empfänger gemessen werden kann. Der Empfänger kann die Werte der Folge a entweder direkt an den Zeitpunkten nT, $n \in \mathbb{Z}$, oder durch Bilden der Skalarprodukte des Signals mit $T\operatorname{sinc}(t/T - n)$ bestimmen. Die Orthogonalität der Basisfunktionen $\operatorname{sinc}(t/T - n)$ des Kanal erlaubt dann eine einfache statistische Untersuchung des Einflusses zufälliger Störungen der Signalfunktion.

b) die Darstellung $f(x) = \sum_{n\in\mathbb{Z}} f(nT)\,\mathrm{sinc}(x/T - n)$.

Beweis: Ist $T > 1/(2\Omega)$, so ist $1/T$ eine Bandbreite zum Intervall $A = [-\Omega, \Omega]$. Desweiteren ist die erzeugenden Funktion durch $\varphi := T\mathcal{F}^{-1}(\chi_A)$,

$$\varphi(x) := T\mathcal{F}^{-1}(\chi_{[-\Omega,\Omega]})(x) = T\frac{\sin(2\pi\Omega x)}{\pi x} = 2\Omega T\,\mathrm{sinc}(2\Omega x)$$

gegeben.

Nach Satz 4.1.4 ist jedes $f \in PW([-\Omega, \Omega])$ durch die Wertefolge $\{f(nT)\}_{n\in\mathbb{Z}}$ vollständig bestimmt und punktweise gegeben durch

$$f(x) = 2\Omega T \sum_{n\in\mathbb{Z}} f(nT)\,\mathrm{sinc}(2\Omega(x - kT))\,.$$

Für $2\Omega T = 1$ ergibt sich der zweite Teil der Behauptung. □

Sei A wieder eine beliebige beschränkte abgeschlossene Teilmenge. Nach den Rechenregeln der Fourier–Transformation ist $PW(A)$ verschiebungsinvariant, mit $f \in PW(A)$ ist auch $\mathcal{T}_a(f) \in PW(A)$, denn $\mathcal{F}(\mathcal{T}_a(f)) = e_a\hat{f}$ hat ebenfalls einen in A enthaltenen Träger. Für die Kardinalreihe der verschobenen Funktion ergibt sich für jeden Parameter $a \in \mathbb{R}$ und für $x \in \mathbb{R}$

$$f(x-a) = (\mathcal{T}_a f)(x) = \sum_{n\in\mathbb{Z}} (\mathcal{T}_a f)(\tfrac{n}{B})\varphi(x - \tfrac{n}{B}) = \sum_{n\in\mathbb{Z}} f(\tfrac{n}{B} - a)\varphi(x - \tfrac{n}{B})$$

Da die Wertefolge zu den Stützstellen Fourier–Koeffizienten zu $\sqrt{B}\hat{f}$ sind, ergibt sich weiter mit der Parsevalschen Gleichung und der Plancherel–Identität

$$\sum_{n\in\mathbb{Z}} |f(\tfrac{n}{B} - a)|^2 = B\|\hat{f}\|^2_{L^2(\mathbb{R})} = B\|f\|^2_{L^2(\mathbb{R})}\,. \tag{4.4}$$

Aus der Verschiebungsformel folgt sofort ein Additionstheorem für φ:

$$\varphi(x-y) = \sum_{n\in\mathbb{Z}} \varphi(\tfrac{n}{B} - y)\varphi(x - \tfrac{n}{B}) = \sum_{n\in\mathbb{Z}} \varphi(x - \tfrac{n}{B})\overline{\varphi(y - \tfrac{n}{B})}\,, \tag{4.5}$$

denn es gilt $\varphi(-z) = \int_A e^{-i2\pi\,\omega z}\,\mathrm{d}\omega = \overline{\varphi(z)}$.

Im allgemeinen Fall wird die erzeugende Funktion φ nicht interpolierend für die Stützstellenmenge $\{\frac{n}{B}\}_{n\in\mathbb{Z}}$ sein, obwohl die Rekonstruktionsformel dies suggeriert. Jedoch nicht für jede Folge $a \in \ell_2(\mathbb{Z})$ ist die Folge der Funktionswerte von $f := \sum_{n\in\mathbb{Z}} a_n\mathcal{T}_{n/B}\varphi$ an den Stützstellen wieder die Folge a.

Weist die Periodisierung $\sum_{n\in\mathbb{Z}} \mathcal{T}_{nB}\chi_A$ der Indikatorfunktion der Menge A überhaupt Nullstellen auf, so ist die Nullstellenmenge $Z := \mathbb{R} \setminus \bigcup_{n\in\mathbb{Z}}(nB + \bar{A})$ offen und B–periodisch. Insbesondere ist $Z \cap [0, B]$ nichtleer und relativ offen. Somit muss auch die Periodisierung

$$g = \sum_{k\in\mathbb{Z}} \mathcal{T}_{kB}\hat{f} = \sum_{n\in\mathbb{Z}} f(\frac{n}{B})e_{-n/B}$$

der Fourier–Transformierten eines jeden $f \in PW(A)$ auf Z verschwinden. Für jedes in der Nullstellenmenge enthaltene Intervall $(a,b) \subset Z$ gilt daher $g(\omega)\chi_{[a,b]} = 0$. Mit der inversen Fourier–Transformation folgt

$$0 = \sum_{n\in\mathbb{Z}} f(\frac{n}{B}) e^{i\pi n \frac{a+b}{B}} \operatorname{sinc}(\tfrac{n(b-a)}{B}) \,,$$

woraus sich lineare Abhängigkeiten zwischen den Gliedern der Wertefolge $\{f(\frac{n}{B})\}_{n\in\mathbb{Z}}$ ableiten lassen. Es kann somit nicht jede Folge aus $\ell_2(\mathbb{C})$ als Wertefolge $\{f(\frac{n}{B})\}_{n\in\mathbb{Z}}$ auftreten. Soll die durch die Wertefolge definierte Abbildung von $PW(A)$ nach $\ell_2(\mathbb{C})$ surjektiv sein, so muss das Auftreten solcher Nullstellenbereiche vermieden werden.

Satz 4.1.6 *Seien $A \subset \mathbb{R}$ eine beschränkte, abgeschlossene Teilmenge und $B > 0$ eine Bandbreite zu A derart, dass $\bigcup_{n\in\mathbb{Z}}(nB + \bar{A}) = \mathbb{R}$ gilt. Sei wieder $\varphi := \frac{1}{B}\mathcal{F}^{-1}(\chi_A)$ die erzeugende Funktion des Frequenzbandes A. Dann gelten die Aussagen*

- *Die Funktionswerte von φ erfüllen $\varphi(\frac{n}{B}) = \delta_{0,n}$ für jedes $n \in \mathbb{Z}$.*
- *$\{\sqrt{B}\mathcal{T}_{\frac{n}{B}}\varphi : \ n \in \mathbb{Z}\}$ ist eine Hilbert–Basis in $PW(A)$.*
- *Die Folgen $\Phi(x) := \{\varphi(x + \frac{m}{B})\}_{m\in\mathbb{Z}} \in \ell_2(\mathbb{Z})$ bilden für jedes $x \in \mathbb{R}$ eine Hilbert–Basis $\{\Phi(x + \frac{n}{B}) : \ n \in \mathbb{Z}\}$ in $\ell_2(\mathbb{Z})$.*

Beweis: Die Periodisierung $g := \sum_{n\in\mathbb{Z}} \mathcal{T}_{nB}\chi_A$ ist, mit Ausnahme einer Menge vom Maß Null, konstant 1. Damit ist g auch fast überall identisch zur Periodisierung der Indikatorfunktion des Intervalls $I := \left[-\frac{B}{2}, \frac{B}{2}\right]$. Die gesuchten Werte von φ bestimmen sich daher zu

$$\begin{aligned}\varphi(\tfrac{n}{B}) =& \frac{1}{B}\int_A \sum_{k\in\mathbb{Z}} e_{n/B}(\omega)\mathcal{T}_{kB}\chi_I \,\mathrm{d}\omega = \frac{1}{B}\sum_{k\in\mathbb{Z}}\int_{kB+A} e_{n/B}(\omega)\chi_I \,\mathrm{d}\omega \\ =& \int_{\mathbb{R}} e_\omega(\tfrac{n}{B})\chi_I \,\mathrm{d}\omega = \operatorname{sinc}(n) = \delta_{0,n} \,.\end{aligned}$$

Somit hat für $a \in \ell_2(\mathbb{Z})$ die Funktion $f \in PW(A)$,

$$x \mapsto f(x) := \sqrt{B}\sum_{n\in\mathbb{Z}} a_n \varphi\left(x - \frac{n}{B}\right)$$

die Funktionswerte $f(\frac{n}{B}) = \sqrt{B}a_n$, $n \in \mathbb{Z}$. Da nun

$$\|f\|_{L^2}^2 = \|\hat{f}\|_{L^2}^2 = \frac{1}{B}\sum_{n\in\mathbb{Z}} |f(\tfrac{n}{B})|^2 = \|a\|_{\ell_2}^2$$

gilt, ist das System $\{\sqrt{B}\mathcal{T}_{\frac{n}{B}}\varphi : \ n \in \mathbb{Z}\}$ ein Orthonormalsystem in $PW(A) \subset L^2(\mathbb{R})$. Da jede Funktion in $PW(A)$ mittels dieses Orthonormalsystems dargestellt werden kann, ist es eine Hilbert–Basis von $PW(A)$.

Nach der Verschiebungsformel gilt für a, f wie eben und beliebiges $x \in \mathbb{R}$ auch

$$\|a\|_{\ell_2}^2 = \|f\|_{L^2}^2 = \frac{1}{B}\sum_{m\in\mathbb{Z}} |f(x + \tfrac{m}{B})|^2 = \left\| \sum_{n\in\mathbb{Z}} a_n \Phi\left(x - \frac{n}{B}\right)\right\|_{\ell_2}^2 .$$

Daher ist auch das System $\{\Phi(x+\frac{n}{B}) : n \in \mathbb{Z}\}$ ein Orthonormalsystem in $\ell_2(\mathbb{Z})$. Für beliebiges $b \in \ell_2(\mathbb{Z})$ sind die Skalarprodukte

$$\left\langle b, \Phi\left(x-\tfrac{n}{B}\right)\right\rangle_{\ell_2} = \sum_{m\in\mathbb{Z}} b_m \overline{\varphi\left(\tfrac{n}{B}-x-\tfrac{m}{B}\right)} = \overline{\tilde{f}(\tfrac{n}{B}-x)}$$

komplex konjugiert zu den Funktionswerten einer Funktion $\tilde{f} \in PW(A)$,

$$z \mapsto \tilde{f}(z) = \sum_{m\in\mathbb{Z}} \overline{b_m}\, \varphi(z-\frac{m}{B})\,.$$

Da auch

$$\|b\|_{\ell_2}^2 = \sum_{n\in\mathbb{Z}} |\tilde{f}(\tfrac{n}{B})|^2 = \sum_{n\in\mathbb{Z}} |\tilde{f}(\tfrac{n}{B}-x)|^2 = \sum_{n\in\mathbb{Z}} \left|\left\langle b, \Phi\left(x-\tfrac{n}{B}\right)\right\rangle_{\ell_2}\right|^2$$

gilt, bildet das System $\{\Phi(x+\frac{n}{B}) : n \in \mathbb{Z}\}$ sogar eine Hilbert–Basis von $\ell_2(\mathbb{C})$. □

Korollar 4.1.7 *Für das Frequenzband $A = I = \left[-\frac{1}{2}, \frac{1}{2}\right]$ mit Bandbreite und $B = 1$ ist die erzeugende Funktion der Kardinalsinus* sinc. *Neben der bekannten Eigenschaft, dass* $\operatorname{sinc}(n) = \delta_{0,n}$ *für $n \in \mathbb{Z}$ gilt, gelten weiter*

- *Das Funktionensystem $S := \{s_n : n \in \mathbb{Z}\} \subset L^2(\mathbb{R})$ mit $s_n := \mathcal{T}^n \operatorname{sinc}$ ist eine Hilbert–Basis von $PW(I)$.*
- *Für beliebige $x, y \in \mathbb{R}$ gilt das Additionstheorem*

$$\sum_{n\in\mathbb{Z}} \operatorname{sinc}(x-n)\operatorname{sinc}(y-n) = \operatorname{sinc}(x-y)\,.$$

- *Für jedes $x \in \mathbb{R}$ ist das System $\{S(x+n) : n \in \mathbb{Z}\}$ eine Hilbert–Basis in $\ell_2(\mathbb{Z})$.*
- *Der Differenzenoperator*

$$S_x(\mathcal{T}) := \sum_{n\in\mathbb{Z}} \operatorname{sinc}(x-n)\mathcal{T}^n : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C}) \tag{4.6}$$

ist unitär. Es gilt $S_y(\mathcal{T}) \circ S_x(\mathcal{T}) = S_{x+y}(\mathcal{T})$, insbesondere ist $S_{-x}(\mathcal{T})$ nicht nur der adjungierte, sondern auch der inverse Operator zu $S_x(\mathcal{T})$.

Beweis: Es verbleibt einzig, die Eigenschaften der Differenzenoperatoren $S_x(\mathcal{T})$ nachzuweisen. Definiert man zu einer gegebenen Folge $a \in \ell_2(\mathbb{C})$ die Kardinalreihe $f := \sum_{n\in\mathbb{Z}} a_n \mathcal{T}_n \operatorname{sinc}$, so besteht die Folge $S_x(\mathcal{T})(a)$ aus den Funktionswerten $\{f(x+n)\}_{n\in\mathbb{Z}}$. Somit erhält $S_x(\mathcal{T})$ die Norm auf $\ell_2(\mathbb{C})$. Weiter folgt mit dem Additionstheorem die Eigenschaft $S_y(\mathcal{T}) \circ S_x(\mathcal{T}) = S_{x+y}(\mathcal{T})$, mit $S_0(\mathcal{T}) = id_{\ell_2(\mathbb{Z})}$ folgt, dass $S_x(\mathcal{T})$ unitär ist. □

4.1.5 Reelle bandbeschränkte Funktionen

Ist f eine reellwertige bandbeschränkte Funktion mit höchster Frequenz Ω, so gilt für die Fourier–Transformierte $\hat{f} = \mathcal{F}(f)$ die Symmetrie

$$\hat{f}(-\omega) = \int_{-\Omega}^{\Omega} f(x) e_\omega(x)\,\mathrm{d}x = \overline{\int_{-\Omega}^{\Omega} f(x)\overline{e_\omega(x)}\,\mathrm{d}x} = \overline{\hat{f}(\omega)}\,.$$

Der Träger von $\hat{f}$ muss also immer symmetrisch zum Nullpunkt sein. Die zum Intervall $[-\Omega, \Omega]$ nächsteinfache Variante für eine zum Nullpunkt symmetrische Menge ist die Vereinigung eines Intervalls mit dem Nullpunkt gespiegelten Intervall. Für $0 < a < b$ hat eine derart symmetrische Menge die Form $[-b, -a] \cup [a, b]$.

Eine solche Menge hat sicher den Wert $B = 2b$ als Bandbreite. Jedoch gibt es Fälle, in welchen eine wesentlich kleinere Bandbreite möglich ist. So kann dem Frequenzband $[-10, -9] \cup [9, 10]$ neben der Bandbreite 20 auch die Bandbreite 2 zugeordnet werden.

Lemma 4.1.8 *Seien $0 < a < b < \infty$ und $A := [-b, -a] \cup [a, b]$. Dann ist jedes*

$$B \in \bigcup_{N \in \mathbb{N}:\, 0 < N \leq \frac{a}{b-a}} \left[\frac{2b}{N+1}, \frac{2a}{N} \right] \cup [2b, \infty)$$

eine Bandbreite von A.

Beweis: Seien $B > 0$ und $n, m \in \mathbb{Z}$. Damit die Mengen $mB + A$ und $nB + A$ bis auf Randpunkte disjunkt sind, dürfen sich insbesondere die beiden Teilintervalle nicht überlappen. Dies ist erfüllt, wenn entweder für die inneren Grenzen die Ungleichung $mB - a \leq nB + a$ oder für die äußeren die Ungleichung $mB - b \geq nB + b$ erfüllt ist. D.h., es muss entweder $B(m - n) \leq 2a$ oder $B(m - n) \geq 2b$ gegeben sein.

Beide Ungleichungen sollen für beliebige Paare $(m, n) \in \mathbb{Z}^2$ erfüllt sein, also auch für beliebige Differenzen $N := m - n \in \mathbb{Z}$. Es gibt immer ein größtes ganzzahliges Vielfaches von B, welches kleiner oder gleich $2a$ ist. Das darauffolgende Vielfache von B muss dann schon größer oder gleich $2b$ sein. Also muss es ein $N \in \mathbb{N}$ geben mit

$$2b \leq B(N + 1) \text{ und } NB \leq 2a \,.$$

Es seien $N \in \mathbb{N}$ und $B > 0$ so gewählt, dass beide Ungleichungen erfüllt sind. Dann gilt auch $B = (N + 1)B - NB \geq 2(b - a)$, daher können sich auch um ganzzahlige Vielfache von B verschobene Kopien jeweils eines der Teilintervalle nicht überlappen. B ist somit eine Bandbreite von A.

Für $N = 0$ ergibt sich $B \in [2b, \infty)$ bei $N = 0$, für $N > 0$ muss $B \in \left[\frac{2b}{N+1}, \frac{2a}{N} \right]$ gelten. Es gibt nur dann Intervalle vom zweiten, endlichen Typ, wenn es wenigstens das Intervall zu $N = 1$ gibt. Dazu muss $b \leq 2a$ gelten. Das Intervall mit der kleinsten Bandbreite ergibt sich durch den größtmöglichen Wert von N. Dieser muss die Bedingung $\frac{b}{N+1} \leq \frac{a}{N}$ erfüllen, ist also die größte natürliche Zahl, die kleiner oder gleich $\frac{a}{b-a}$ ist. □

Für jedes symmetrische Paar von Intervallen kann nach dem allgemeinen Abtasttheorem die erzeugende Funktion bestimmt werden, die Orthogonalitätseigenschaften gelten nur dann, wenn die Bandbreite und die Gesamtlänge $2(b - a)$ der Intervalle übereinstimmen.

Satz 4.1.9 *Für $0 \leq a < b < \infty$ ist die erzeugende Funktion $\varphi \in PW([-b, -a] \cup [a, b])$ durch*

$$x \mapsto \varphi(x) = 2b \operatorname{sinc}(2bx) - 2a \operatorname{sinc}(2ax)$$

gegeben. Diese erzeugt nur dann ein Orthonormalsystem von $PW([-b,-a]\cup[a,b])$*, wenn es ein* $N\in\mathbb{N}$ *und* $B>0$ *gibt mit* $a=N\frac{B}{2}$ *und* $b=(N+1)\frac{B}{2}$ *und wenn* B *als Bandbreite gewählt wird.*

Beweis: Für die an Null symmetrische Menge $[-b,-a]\cup[a,b]$ erhalten wir die erzeugende Funktion

$$\varphi(x)=\int_a^b (e_x(\omega)+e_x(-\omega))\,\mathrm{d}\omega=\frac{\sin(2\pi bx)-\sin(2\pi ax)}{\pi x}$$
$$=2b\,\mathrm{sinc}(2bx)-2a\,\mathrm{sinc}(2ax)\,.$$

Die Norm $\|\varphi\|_{L^2}=2(b-a)$ ist die einzig mögliche Bandbreite, welche mit φ ein Orthonormalsystem erzeugen kann. $B:=2(b-a)$ ist gleichzeitig die kleinstmögliche Bandbreite, es muss nach Lemma 4.1.8 ein $N\in\mathbb{N}$ geben mit $2b=(N+1)B$ und damit $NB=2b-2(b-a)=2a$. Man überzeugt sich leicht, dass Verschiebungen von $A=[-(N+1)\frac{B}{2},-N\frac{B}{2}]\cup[N\frac{B}{2},(N+1)\frac{B}{2}]$ um ganzzahlige Vielfache von B ganz $\mathbb{R}$ überdecken. □

4.2 Approximation in verschiebungsinvarianten Teilräumen

Dass von einem physikalischen Prozess einzelne Werte einer Größe zu einem bestimmten Zeitpunkt bestimmt werden können, ist eine sehr starke Idealisierung eines realen Messvorgangs. Jede reale Messung wird ein Mittelwert aus Werten der zu messenden Größe über ein Zeitintervall sein.

Um dies zu modellieren, kann die zu messende, zeitveränderliche Größe als Funktion $f\in L^2(\mathbb{R})$ angenommen werden. Meist wird noch die Stetigkeit der Funktion f vorausgesetzt, so dass von einzelnen Funktionswerten gesprochen werden kann. Dem Messvorgang entspricht dann das Skalarprodukt von f mit einer weiteren Funktion, die das Messgerät modelliert. Die Schätzung bzw. näherungsweise Rekonstruktion des realen Verlaufs der gemessenen Größe aus einer Zeitreihe von Messwerten, d.h. einer Abtastfolge, muss dann diesem Messvorgang angepasst werden.

Um die Güte des Paars aus Abtastung und Rekonstruktion zu beurteilen, betrachtet man deren Wirkung auf sich sehr langsam ändernde Funktionen, d.h. bandbeschränkte Funktionen mit sehr kleiner Maximalfrequenz. Dies legt nahe, beide Vorgänge in ihrer Wirkung auf die Fourier–Transformierte der zu messenden Funktion bzw. der rekonstruierten Funktion zu betrachten.

Die nachfolgend dargestellten Zusammenhänge wurden in [SF73] erstmals im Zusammenhang mit der Methode der finiten Elemente untersucht.

4.2.1 Von der Abtastfolge zum Signal

Es sei eine Abtastfolge $c=\{c_n\}_{n\in\mathbb{Z}}$ zu einer Folge von äquidistanten Messzeitpunkten. Es sei zunächst angenommen, dass das Intervall zwischen zwei Messungen die Länge 1 hat und die Messzeitpunkte den ganzen Zahlen zugeordnet werden können. Die Rekonstruktion erfolge

entsprechend der ebenfalls vorauszusetzenden verschiebungsinvarianten Natur der Messung durch eine Linearkombination von ganzzahligen Verschiebungen einer Funktion $\varphi \in L^2(\mathbb{R})$, d.h. durch deren *Synthese–Operator* als $\mathcal{E}_\varphi(c) := \sum_{n\in\mathbb{Z}} c_n \mathcal{T}_n \varphi$.

Diese Konstruktion ist nur sinnvoll, wenn die Reihe für alle $c \in \ell_2(\mathbb{C})$ in $L^2(\mathbb{R})$ konvergiert. Dazu muss φ ein Bessel–System erzeugen (vgl. Definition C.2.1 auf Seite 220), d.h. es muss eine Konstante $B > 0$ geben, so dass

$$\sum_{n\in\mathbb{Z}} |\hat{\varphi}(\omega + n)|^2 \leq B$$

fast überall gibt. Dann ist für jedes $c \in \ell_2(\mathbb{Z})$ mit Fourier–Reihe $\hat{c} := \sum_{n\in\mathbb{Z}} c_n e_{-n}$ die Funktionenreihe $\mathcal{E}_\varphi(c)$ quadratintegrabel, genauer gilt unter Benutzung der Plancherel–Identität

$$\begin{aligned} \|\mathcal{E}_\varphi(c)\|_{L^2}^2 &= \left\| \sum_{n\in\mathbb{Z}} c_n e_{-n} \hat{\varphi} \right\|_{L^2}^2 \\ &= \int_0^1 |\hat{c}(\omega)|^2 \left| \sum_{n\in\mathbb{Z}} \hat{\varphi}(\omega + n) \right|^2 \mathsf{d}\omega \leq B\|c\|_{\ell_2}^2 . \end{aligned}$$

Der lineare Operator $\mathcal{E}_\varphi : \ell_2(\mathbb{C}) \to L^2(\mathbb{R})$ ist somit beschränkt mit Schranke $\sqrt{B}$. Ein anderer Ausdruck dafür ist, dass die *Prä–Gramsche Faser* $J_\varphi : \mathbb{R} \to \ell_2(\mathbb{C})$, $\omega \mapsto J_\varphi(\omega) := \{\hat{\varphi}(\omega + n)\}_{n\in\mathbb{Z}}$ (s. Anhang C.2, Definition C.2.3) in die Kugel mit Radius $\sqrt{B}$ in $\ell_2(\mathbb{C})$ abbildet. Die Gramsche Faser der rekonstruierten Funktion ergibt sich zu

$$J_{\mathcal{E}_\varphi(c)}(\omega) = \hat{c}(\omega) J_\varphi(\omega) .$$

4.2.2 Vom Signal zur Abtastfolge

Im $L^2(\mathbb{R})$ werden das Bilden eines Mittelwerts über einem beschränkten Intervall oder komplexere Bildungsvorschriften durch das Skalarprodukt von f mit einer Funktion $\tilde{\varphi} \in L^2(\mathbb{R})$ ausgedrückt.

Nehmen wir wieder an, dass die Zeitskala so gewählt wurde, dass die Zeitpunkte der einzelnen Messungen den ganzen Zahlen entsprechen. Dann entspricht das Aufstellen der Messreihe der Anwendung des Analyse–Operators zu $\tilde{\varphi}$, d.h. des adjungierten Operators zum Synthese–Operator, $\mathcal{E}_{\tilde{\varphi}}^*(f) = \{\langle f, \mathcal{T}^n \tilde{\varphi} \rangle\}_{n\in\mathbb{Z}}$.

Damit die so erhaltene Folge für alle $f \in L^2(\mathbb{R})$ im Folgenraum $\ell_2(\mathbb{C})$ zu liegen kommt, müssen wir wieder voraussetzen, dass die Funktion $\tilde{\varphi} \in L^2(\mathbb{R})$ ein *Bessel–System* erzeugt, es also eine Konstante $\tilde{B} > 0$ gibt, so dass

$$\sum_{n\in\mathbb{Z}} \left|\hat{\tilde{\varphi}}(\omega + n)\right|^2 \leq \tilde{B}$$

fast überall gilt (vgl. Definition C.2.1). Dann ist die Folge der Skalarprodukte quadratsummierbar, also $\mathcal{E}_{\tilde{\varphi}}^*(f) \in \ell_2(\mathbb{Z})$. Denn nach obigem ist der Synthese–Operator zu $\tilde{\varphi}$ beschränkt, somit

ist nach Lemma A.2.4 auch dessen adjungierter Analyse–Operator beschränkt mit derselben Schranke $\sqrt{\tilde{B}}$,

$$\|\mathcal{E}_{\tilde{\varphi}}^{*}(f)\|_{\ell_2}^2 \leq \tilde{B}\|f\|_{L^2}^2 .$$

Die Fourier–Reihe $\hat{c}$ der Abtastfolge $c := \mathcal{E}_{\tilde{\varphi}}^{*}(f)$ ergibt sich mittels der Prä–Gramschen Fasern zu

$$\hat{c}(\omega) = \left\langle J_f(\omega),\, J_{\tilde{\varphi}}(\omega) \right\rangle_{\ell_2(\mathbb{C})} .$$

4.2.3 Vom Signal zur Rekonstruktion

Seien B und $\tilde{B}$ die Bessel–Konstanten zu φ und $\tilde{\varphi}$. Dann ist die Verknüpfung $P_{\tilde{\varphi},\varphi} := \mathcal{E}_{\varphi} \circ \mathcal{E}_{\tilde{\varphi}}^{*}$ von Abtastung und Rekonstruktion ein beschränkter linearer Operator mit einer Schranke $\sqrt{B\tilde{B}}$. Wir nennen diesen im Folgenden *Transferoperator* zum Paar $(\tilde{\varphi}, \varphi)$. Dieser Operator erzeugt aus der Ausgangsfunktion die rekonstruierte Funktion und hat für jedes $f \in L^2(\mathbb{R})$ die Darstellung

$$P_{\tilde{\varphi},\varphi}(f) = \sum_{n\in\mathbb{Z}} \langle\, f,\, \mathcal{T}^n\tilde{\varphi} \,\rangle\, \mathcal{T}^n\varphi \,. \tag{4.7a}$$

Der Transferoperator kann auch mittels der Prä–Gramschen Faser (s. Anhang C.2) ausgedrückt werden. Für fast jedes $\omega \in \mathbb{R}$ gilt

$$J_{P_{\tilde{\varphi},\varphi}(f)}(\omega) = \left\langle J_f(\omega),\, J_{\tilde{\varphi}}(\omega) \right\rangle_{\ell_2} \cdot J_{\varphi}(\omega) \,. \tag{4.7b}$$

Wir können die Genauigkeit der durch den Transferoperator erzeugten Approximation auf ausgewählten Unterräumen des Funktionenraums $L^2(\mathbb{R})$ testen. Da die Untersuchung mittels der Fourier–Transformation erfolgt, bieten sich die Unterräume bandbeschränkter Funktionen an. Wir interessieren uns also für Schranken E_Ω, $\Omega > 0$, so dass für alle Funktionen $f \in PW([-\Omega,\Omega])$ die Abschätzung

$$\|f - P_{\tilde{\varphi},\varphi}(f)\| \leq E_\Omega\|f\|$$

für den Approximationsfehler gilt.

Betrachten wir nun eine bandbeschränkte Funktion $f \in PW([-\Omega,\Omega])$ mit höchster Frequenz $\Omega \in \left(0, \frac{1}{2}\right)$. In jeder Prä–Gramschen Faser $J_f(\omega) \in \ell_2(\mathbb{Z})$ einer solchen Funktion verschwinden alle Glieder bis auf Ausnahme eines einzigen, insbesondere für $\omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ ist höchstens das Glied mit Index Null von Null verschieden. Somit ist $J_f(\omega)$ proportional zur Einheitsfolge $\delta_0 = \{\delta_{0,n}\}$, definiert durch das Kronecker-Delta als $\delta_{0,0} = 1$ und $\delta_{0,n} = 0$ für $n \neq 0$; und es gilt $J_f(\omega) = \hat{f}(\omega)\delta_0$. Mit Gleichung (4.7b) gilt also

$$\begin{aligned}
\|P_{\tilde{\varphi},\varphi}(f) - f\|_{L^2}^2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \|J_{Pf}(\omega) - J_f(\omega)\|_{\ell_2}^2 \,\mathrm{d}\omega \\
&= \int_{-\Omega}^{\Omega} \left\| \left\langle \hat{f}(\omega)\delta_0,\, J_{\tilde{\varphi}}(\omega) \right\rangle_{\ell_2} J_{\varphi}(\omega) - \hat{f}(\omega)\delta_0 \right\|_{\ell_2}^2 \mathrm{d}\omega \\
&= \int_{-\Omega}^{\Omega} |\hat{f}(\omega)|^2 \left\| \overline{\hat{\tilde{\varphi}}(\omega)}\, J_{\varphi}(\omega) - \delta_0 \right\|_{\ell_2}^2 \mathrm{d}\omega
\end{aligned} \tag{4.8}$$

Wir erhalten somit E_Ω als Supremum der Einzelfehler

$$e(\omega) := \sqrt{\sum_{n\in\mathbb{Z}} \left| \overline{\hat{\tilde{\varphi}}(\omega)}\, \hat{\varphi}(\omega+n) - \delta_{0,n} \right|^2}$$

für $\omega \in [-\Omega, \Omega]$.

$$\|P_{\tilde{\varphi},\varphi}(f) - f\|_{L^2}^2 \leq \sup_{\omega\in[-\Omega,\Omega]} e(\omega)^2 \int_{-\Omega}^{\Omega} |\hat{f}(\omega)|^2 \,\mathrm{d}\omega = E_\Omega^2 \|f\|_{L^2}^2 \tag{4.9}$$

Unter dem Gesichtspunkt der Approximation ist zu fordern, dass für Ω nahe Null der Approximationsfehler E_Ω klein wird, also E_Ω als Funktion in Ω im Nullpunkt stetig mit Wert Null ist. Nach der Konstruktion von E_Ω ist dies nun auch für die Einzelfehler $e(\omega)$ zu fordern, insbesondere kann man verlangen, dass für ω nahe Null die Werte der Fourier–Transformierten $\hat{\tilde{\varphi}}(\omega)$ nahe 1 und die Prä–Gramsche Faser $J_\varphi(\omega) \in \ell_2(\mathbb{Z})$ nahe der Einheitsfolge δ_0 liegen. Dies ist auch, nach Austausch eines konstanten Faktors, der allgemeine Fall.

Diese Forderung kann man in verschiedenen Stufen exakt fassen. Wir können fordern, dass die Prä–Gramschen Fasern J_φ und $J_{\tilde{\varphi}}$ im Nullpunkt gegen die Einheitsfolge δ_0 konvergieren. Wir können weiter Forderungen an die Geschwindigkeit dieser Konvergenz stellen.

4.2.4 Landau–Symbole

Definition 4.2.1 (Landau–Symbole) *Seien $\mathcal{B}$ ein $\mathbb{C}$–Banach–Raum, $\varepsilon > 0$ und $f, g : [-\varepsilon, \varepsilon] \to \mathcal{B}$ eine Funktion. Wir notieren*

- *$f = g + O_\mathcal{B}(x^r)$, wenn es ein $r \in \mathbb{N}$ und ein $C > 0$ derart gibt, dass $\|f(x) - g(x)\|_\mathcal{B} \leq C\,|x|^r$ für alle $x \in [-\varepsilon, \varepsilon]$ gilt.*
- *$f = g + o_\mathcal{B}(x^r)$, wenn sowohl f als auch g beschränkt sind und es ein $r \in \mathbb{N}$ derart gibt, dass $\lim_{x\to 0} |x|^{-r} \|f(x) - g(x)\|_\mathcal{B} = 0$ gilt.*

Somit bedeuten $f = f(0) + o_\mathcal{B}(1)$ oder $f = f(0) + O_\mathcal{B}(x)$, dass f im Nullpunkt stetig ist; $f = O_\mathcal{B}(1)$ ist äquivalent dazu, dass f beschränkt ist. Es gelten die Rechenregeln $o_\mathcal{B}(x^r) + o_\mathcal{B}(x^s) = o_\mathcal{B}(x^{\min(r,s)})$ und $O_\mathbb{C}(x^r)\, o_\mathcal{B}(x^s) = o_\mathcal{B}(x^{r+s})$. Dass $f = g + o_\mathcal{B}(x^r)$ für zwei beschränkte Funktionen $f, g : [-\varepsilon, \varepsilon] \to \mathcal{B}$ gilt, ist nach Definition äquivalent dazu, dass $\|f - g\|_\mathcal{B} = o_\mathbb{C}(x^r)$ gilt. Aus $f = o_\mathcal{B}(x^r)$ folgt $f = O_\mathcal{B}(x^r)$, aus $f = O_\mathcal{B}(x^r)$ wiederum folgt $f = o_\mathcal{B}(x^{r-1})$.

Lemma 4.2.2 *Seien $\mathcal{H}$ ein $\mathbb{C}$–Hilbert–Raum, $\varepsilon > 0$, $r \in \mathbb{N}$ und $f_1, f_2, g_1, g_2 : [-\varepsilon, \varepsilon] \to \mathcal{H}$ Funktionen, für welche $f_k = g_k + o_\mathcal{H}(x^r)$, $k = 1, 2$, gilt. Dann ist*

$$\langle f_1,\, f_2 \rangle_\mathcal{H} = \langle g_1(x),\, g_2(x) \rangle_\mathcal{H} + o_\mathbb{C}(x^r)\,.$$

Beweis: Nach Definition sind alle Funktionen beschränkt, insbesondere $f_2 = O_\mathcal{H}(1)$, $g_1 = O_\mathcal{H}(1)$, und damit

$$\begin{aligned} |\langle f_1,\, f_2 \rangle_\mathcal{H} - \langle g_1,\, g_2 \rangle_\mathcal{H}| &= |\langle f_1 - g_1,\, f_2 \rangle_\mathcal{H} - \langle g_1,\, f_2 - g_2 \rangle_\mathcal{H}| \\ &\leq \|f_1 - g_1\|_\mathcal{H}\, \|f_2\|_\mathcal{H} + \|g_1\|_\mathcal{H}\, \|f_2 - g_2\|_\mathcal{H} = o_\mathbb{C}(x^r) \end{aligned}$$

□

Lemma 4.2.3 *Seien $\varepsilon > 0$ und $f : [-\varepsilon, \varepsilon] \to \mathbb{C}$ eine r–fach stetig differenzierbare Funktion mit $f(0) = 1$. Dann gibt es ein Polynom $g \in \mathbb{C}[X]$ vom Grad höchstens r, so dass für deren Produkt $g\,f - 1 = o_{\mathbb{C}}(x^r)$ gilt.*

Beweis: Sei p das Taylor–Polynom von f im Nullpunkt zum Grad r. Dann gibt es für jedes $x \in [-\varepsilon, \varepsilon]$ ein $\theta \in (0,1)$ mit

$$f(x) = 1 + f'(0)x + \cdots + f^{(r-1)}(0)\frac{x^{r-1}}{(r-1)!} + f^{(r)}(\theta x)\frac{x^r}{r!} = p(x) + \frac{f^{(r)}(\theta x) - f^{(r)}(0)}{r!}x^r\,.$$

Da die r–te Ableitung stetig vorausgesetzt ist, gilt somit $f(x) = p(x) + o_{\mathbb{C}}(x^r)$. Wegen $p(0) = 1$ gibt es eine Potenzreihe $q \in \mathbb{C}[[X]]$, welche in einer Umgebung der Null konvergiert und zu p reziprok ist. Sei $g := 1 + q_1 X + \cdots + q_r X^r$ der Anfang dieser Potenzreihe. Dabei bezeichne $X : \mathbb{R} \to \mathbb{C}$ ebenfalls die identische Funktion, $X(x) := x$. Somit gilt $g\,p = 1 + O_{\mathbb{C}}(x^{r+1})$, also auch $g\,f = 1 + o_{\mathbb{C}}(x^r)$. □

Man kann zu jedem Polynom ein trigonometrisches Polynom mit derselben Taylor–Reihe im Nullpunkt finden. Zusammen mit obigem Resultat ergibt sich folgendes Lemma.

Lemma 4.2.4 *Seien $\varepsilon > 0$ und $f : [-\varepsilon, \varepsilon] \to \mathbb{C}$ eine r–fach stetig differenzierbare Funktion mit $f(0) = 1$. Dann gibt es eine endliche Folge $a \in \ell_{\mathrm{fin}}(\mathbb{Z})$, so dass mit deren Fourier–Reihe $\hat{a}f = 1 + o_{\mathbb{C}}(x^r)$ gilt.*

Beweis: Sei das Polynom $g = 1 + q_1 X + \cdots + q_r X^r$ vom Grad r wie oben konstruiert. Sei $K \subset \mathbb{Z}$ eine beliebige Teilmenge mit $r+1$ Elementen. Für jedes $k \in K$ ist $e_{-k} = 1 + \sum_{n=1}^{r} \frac{(-i2\pi k)^n}{n!} X^n + O(x^{r+1})$. Es muss also eine Linearkombination $\sum_{k\in K} a_k e_{-k}$ dieser Funktionen gefunden werden, deren Koeffizienten das lineare Gleichungssystem

$$\sum_{k\in K} a_k = 1, \quad \sum_{k\in K} k^n a_k = \tfrac{n!}{(-i2\pi)^n}\, q_n, \quad n = 1, \ldots, r$$

erfüllen. Da die Determinante der Systemmatrix eine Vandermondesche ist, ist dieses Gleichungssystem immer lösbar. □

4.2.5 Approximation mittels gestauchtem Transferoperator

Definition 4.2.5 *Seien $\varphi \in L^2(\mathbb{R})$ und $r \in \mathbb{N}$. φ hat eine* Approximationsordnung *r, wenn es ein $c \in \ell_1(\mathbb{Z})$ mit $\hat{c}(0) = 1$ gibt, so dass für die Prä–Gramsche Faser $J_\varphi : \left[-\frac{1}{2}, \frac{1}{2}\right] \to \ell_2(\mathbb{Z})$*

$$J_\varphi = \hat{c}\delta_0 + o_{\ell_2}(\omega^r)$$

gilt. Wir sagen dann auch, dass φ eine Approximationsbedingung (der Ordnung $r \in \mathbb{N}$) erfüllt.

Als direkte Konsequenz dieser Eigenschaft erhalten wir, dass $\mathcal{T}^n\hat{\varphi} = o_{\mathbb{C}}(\omega^r)$ für jedes $n \in \mathbb{Z} \setminus \{0\}$ und $\hat{\varphi} = \hat{c} + o_{\mathbb{C}}(\omega^r)$ gilt.

Lemma 4.2.6 *Erfüllen $\varphi, \tilde{\varphi} \in L^2(\mathbb{R})$ eine Approximationsbedingung der Ordnung r, so erzeugen beide Funktionen jeweils ein verschiebungsinvariantes Bessel–System. Sind $c, \tilde{c} \in \ell_1(\mathbb{Z})$ Folgen, mit welchen $J_\varphi = \hat{c}\delta_0 + o_{\ell_2}(\omega^r)$ und $J_{\tilde{\varphi}} = \hat{\tilde{c}}\delta_0 + o_{\ell_2}(\omega^r)$ gelten, und gilt weiterhin $\hat{c}\,\overline{\hat{\tilde{c}}} = 1 + o_{\mathbb{C}}(\omega^r)$, so folgt*

$$\overline{\hat{\varphi}(\omega)}\, J_\varphi(\omega) = \delta_0 + o_{\ell_2}(\omega^r)\,.$$

Beweis: Nach Definition der Landau–Notation folgt aus der Voraussetzung, dass die Gramschen Fasern J_φ und $J_{\tilde{\varphi}}$ beschränkt sind. Somit sind die von φ bzw. $\tilde{\varphi}$ erzeugten verschiebungsinvarianten Systeme Bessel–Systeme.

Gilt $J_{\tilde{\varphi}} = \hat{\tilde{c}}\delta_0 + o_{\ell_2}(\omega^r)$, so auch $\hat{\tilde{\varphi}}(\omega) = \hat{\tilde{c}} + o_{\mathbb{C}}(\omega^r)$, mit der zweiten Beziehung $J_\varphi = \hat{c}\delta_0 + o_{\ell_2}(\omega^r)$ also

$$\overline{\hat{\tilde{\varphi}}}\, J_\varphi = \overline{\hat{\tilde{c}}}\,\hat{c}\delta_0 + o_{\ell_2}(\omega^r) = \delta_0 + o_{\ell_2}(\omega^r)$$

□

Jedem Transferoperator $P_{\tilde{\varphi},\varphi}$ können wir eine Familie $\{P_S : S \in \mathbb{R}_+\}$ dilatierter Operatoren zuordnen, indem wir zunächst die zu transformierende Funktion strecken, dann abtasten und rekonstruieren und danach wieder stauchen. D.h. für einen gegebenen Faktor $S > 1$ transformieren wir eine Funktion $f \in L^2(\mathbb{R})$ zu

$$P_S(f) := \mathcal{D}_S \circ P_{\tilde{\varphi},\varphi} \circ \mathcal{D}_S^{-1}(f)\,.$$

Die so entstehende Funktionenfamilie $\{P_S(f) : S > 1\}$ soll nun daraufhin untersucht werden, ob sie eine Approximation von f darstellt, d.h. ob $\lim_{S\to\infty} P_S(f) = f$ für die Konvergenz in $L^2(\mathbb{R})$ gilt.

Sei dazu f zunächst als bandbeschränkt gewählt, $f \in PW([-\Omega,\Omega])$ für ein $\Omega > 0$. Nach den Rechenregeln (B.4) gilt $\mathcal{D}_S^{-1} f \in PW([-S^{-1}\Omega, S^{-1}\Omega])$ für jedes $S > 0$. Für $S > 2\Omega$ gilt also $\mathcal{D}_S^{-1} f \in PW([-\frac{1}{2},\frac{1}{2}])$. Unter den Voraussetzungen des vorhergehenden Lemmas 4.2.6 und nach Ungleichung (4.9) können wir den Abstand von $P_S(f) = (\mathcal{D}_S P_{\varphi,\tilde{\varphi}} \mathcal{D}_S^{-1})(f)$ zu f durch eine Potenz von S^{-1} abschätzen, wenn f genügend oft differenzierbar und S groß genug ist.

Die Beispielklasse der bandbeschränkten Funktionen ist beliebig oft stetig diffenzierbar. Für jedes $f \in PW([-\Omega,\Omega]), \Omega > 0$ liegen die Ableitungen von f sämtlich wieder in $PW([-\Omega,\Omega])$, denn es gilt

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = \frac{\mathrm{d}}{\mathrm{d}x} \int_{-\Omega}^{\Omega} \hat{f}(\omega) e^{i2\pi\omega x}\,\mathrm{d}\omega = \int_{-\Omega}^{\Omega} (i2\pi\omega)\hat{f}(\omega) e^{i2\pi\omega x}\,\mathrm{d}\omega$$

und $p\hat{f} \in L^2([-\Omega,\Omega])$ für jedes Polynom $p \in \mathbb{C}[\omega]$.

Lemma 4.2.7 *Es seien $\varphi, \tilde{\varphi} \in L^2(\mathbb{R})$ so gegeben, dass deren Prä–Gramsche Fasern $J_\varphi, J_{\tilde{\varphi}} : \left[-\frac{1}{2},\frac{1}{2}\right] \to \ell_2(\mathbb{Z})$ beschränkt sind und $\overline{\hat{\varphi}(\omega)}\, J_\varphi(\omega) = \delta_0 + o_{\ell_2}(\omega^r)$ gilt. Sei $\{P_S : S > 0\}$, die Familie der gestauchten Transferoperatoren $P_S := \mathcal{D}_S P_{\tilde{\varphi},\varphi} \mathcal{D}_S^{-1}$.*

Dann gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass für jede bandbeschränkte Funktion $f \in PW([-\Omega, \Omega])$ mit höchster Frequenz $\Omega > 0$ und jedes $S \geq \delta^{-1}\Omega$ gilt:

$$\|P_S(f) - f\|_{L^2} \leq \varepsilon S^{-r} \|f^{(r)}\|_{L^2} .$$

Beweis: Nach Definition der Landau–Notation gibt es für jedes ε ein $\delta > 0$, so dass für alle $\omega \in [-\delta, \delta]$

$$e(\omega) = \left\| \overline{\hat{\tilde{\varphi}}(\omega)}\, J_\varphi(\omega) - \delta_0 \right\|_{\ell_2} \leq \varepsilon |2\pi\omega|^r$$

gilt.

Sind $f \in PW([-\Omega, \Omega])$ und $S \in \mathbb{R}$ mit $\delta S \geq \Omega$ gegeben, so ist $g := \mathcal{D}_S^{-1} f \in PW([-\delta, \delta])$, somit gilt nach Ungleichung (4.8)

$$\left\| P_{\varphi,\tilde{\varphi}} g - g \right\|_{L^2}^2 \leq \varepsilon^2 \int_{-\delta}^{\delta} |(i2\pi\omega)^r \hat{g}(\omega)|^2 \, \mathrm{d}\omega = \varepsilon^2 \|g^{(r)}\|_{L^2}^2 .$$

Nach der Kettenregel der Differentiation gilt $g^{(r)} = S^{-r} \mathcal{D}_S^{-1} f^{(r)}$, und da Dilatation mit Faktor S die L^2–Norm lediglich um den konstanten Faktor $\sqrt{S}$ ändert, erhalten wir

$$\|P_S f - f\|_{L^2} = \left\| \mathcal{D}_S P_{\varphi,\tilde{\varphi}} g - \mathcal{D}_S g \right\|_{L^2} \leq \varepsilon \|\mathcal{D}_S g^{(r)}\|_{L^2} = \varepsilon S^{-r} \|f^{(r)}\|_{L^2} .$$

□

Beispiel: Wir betrachten Abtastung und Rekonstruktion mit der Rechteckfunktion χ_I über dem Intervall $I = \left[-\frac{1}{2}, \frac{1}{2}\right]$, d.h. den Transfer–Operator zu $\varphi = \tilde{\varphi} = \chi_I$. Dieser ist auch der Projektor auf den Unterraum V_0 der Multiskalenanalyse der Haar–Wavelets, s. Abschnitt 5.2.

Es gilt $\hat{\varphi} = \hat{\tilde{\varphi}} = \mathrm{sinc}$ und damit $J_\varphi(\omega) = J_{\tilde{\varphi}}(\omega) = S(\omega) := \{\mathrm{sinc}(n + \omega)\}_{n \in \mathbb{Z}}$. Damit haben diese Prä–Gramschen Fasern alle die Länge 1 und sind stetig in allen $\omega \in \mathbb{R}$. Im Nullpunkt gilt $S(0) = \delta_0$ und es gilt für die Abschätzung der Konvergenzgeschwindigkeit

$$\begin{aligned} e(\omega)^2 = \left\| \overline{\hat{\tilde{\varphi}}(\omega)}\, J_\varphi(\omega) - \delta_0 \right\|_{\ell_2}^2 &= \mathrm{sinc}(\omega)^2 - 2\,\mathrm{sinc}(\omega)^2 + 1 = 1 - \mathrm{sinc}(\omega)^2 \\ &= 8(\pi\omega)^2 \sum_{k=0}^{\infty} \frac{(-1)^k (2\pi\omega)^{2k}}{(2k+4)!} \leq \frac{(\pi\omega)^2}{3} = O_{\mathbb{R}}(\omega^2) . \end{aligned}$$

Da $e(\omega)$ im Nullpunkt linear in $|\omega|$ ist, hat χ_I die Approximationsordnung 0.

Desweiteren ist $e(\omega)$ für $\omega \in I$ symmetrisch und konvex mit Minimum im Nullpunkt. Es gilt also $E_\Omega = e(\Omega)$ und für jedes $f \in PW([-\Omega, \Omega])$ mit $\Omega \in \left(0, \frac{1}{2}\right)$ erhalten wir als Schranke des Approximationsfehlers

$$\|Pf - f\|_{L^2} \leq \sqrt{1 - \mathrm{sinc}(\Omega)^2} \|f\|_{L^2} .$$

Für den relativen Fehler erhalten wir beispielsweise $\sqrt{1 - \mathrm{sinc}(\frac{2}{37})^2} < \frac{1}{10}$ und $\sqrt{1 - \mathrm{sinc}(\frac{1}{182})^2} < \frac{1}{100}$. Es sei daran erinnert, dass nach dem WKS–Abtasttheorem 4.1.5 Funktionen in $PW([-\frac{1}{182}, \frac{1}{182}])$ durch ihre Funktionswerte an den Stellen $\{91\, n\}_{n \in \mathbb{Z}}$ eindeutig bestimmt sind. Um einen relativen Fehler kleiner als 1% zu sichern, müssen bei der Abtastung mittels χ_I also 90 mal mehr Abtastwerte bei gleichem Zeitintervall bestimmt werden.

Andererseits ergibt sich bei Abtastung mittels $\varphi = \tilde{\varphi} = \mathrm{sinc}$ für beliebige $f \in PW(I)$ der Fehler Null, da der zugehörige Transfer–Operator auf $PW(I)$ die Identität ist. Dies ergibt sich ebenfalls aus der Ungleichung (4.8), denn es gilt für jedes $\omega \in I$ sowohl $\hat{\tilde{\varphi}}(\omega) = \chi_I(\omega) = 1$ als auch $J_\varphi(\omega) = \delta_0$, somit $\overline{\hat{\tilde{\varphi}}}\, J_\varphi - \delta_0 = 0$.

Der Zusammenhang zwischen Ableitung und Fourier–Transformation erlaubt es, Klassen differenzierbarer Funktionen in $L^2(\mathbb{R})$ zu identifizieren.

Definition 4.2.8 *Der* Sobolew–Raum *der r–fach in $L^2(\mathbb{R})$ differenzierbaren Funktionen ist definiert als*

$$\mathcal{H}^r(\mathbb{R}) := \left\{ f \in L^2(\mathbb{R}) : \int_{\mathbb{R}} (1+|\omega|^2)^r |\mathcal{F}(f)(\omega)|^2 \,\mathrm{d}\omega < \infty \right\} .$$

Ist $f \in \mathcal{H}^r(\mathbb{R})$, so sind die Ableitungen von f definiert als $f^{(k)} = \mathcal{F}^*\Big((i2\pi\omega)^k \mathcal{F}(f)\Big)$ für alle $k = 0, \ldots, r$.

Satz 4.2.9 (Überabtasttheorem) *Seien $\varphi, \tilde{\varphi} \in L^2(\mathbb{R})$ derart gewählt, dass sie eine Approximationsordnung $r \in \mathbb{N}$ haben und die Folgen $c, \tilde{c} \in \ell_1(\mathbb{Z})$, mit welchen die Approximationsbedingungen $J_\varphi = \hat{c}\delta_0 + o_{\ell_2}(\omega^r)$ bzw. $J_{\tilde{\varphi}} = \hat{\tilde{c}}\delta_0 + o_{\ell_2}(\omega^r)$ gelten, zusätzlich $\overline{\hat{\tilde{c}}}\,\hat{c} = 1 + o_{\mathbb{C}}(\omega^r)$ erfüllen. Sei $\{P_S : S > 0\}$, die Familie der gestauchten Transferoperatoren $P_S := \mathcal{D}_S P_{\tilde{\varphi},\varphi} \mathcal{D}_S^{-1}$.*

Dann gilt für jedes $f \in \mathcal{H}^r(\mathbb{R})$

$$\lim_{S\to\infty} S^r \|f - P_S(f)\|_{L^2(\mathbb{R})} = 0 .$$

Beweis: Sei für jedes $R > 0$ mit $\pi_R : L^2(\mathbb{R}) \to PW([-R,R])$ die orthogonale Projektion auf den Unterraum bandbeschränkter Funktionen mit höchster Frequenz R bezeichnet, $f \mapsto \pi_R(f) := \mathcal{F}^*(\chi_{[-R,R]}\mathcal{F}(f))$. Nach den Lemmata 4.2.6 und 4.2.7 gibt es für jedes $\varepsilon > 0$ ein $\delta > 0$, so dass für jedes $f \in \mathcal{H}^r(\mathbb{R})$ und jedes $S > 0$

$$S^r \left\| P_S(\pi_{\delta S}(f)) - \pi_{\delta S}(f) \right\|_{L^2} \le \varepsilon \left\| \left(\tfrac{\mathrm{d}}{\mathrm{d}x}\right)^r \pi_{\delta S}(f) \right\|_{L^2} = \varepsilon \left\| \pi_{\delta S}(f^{(r)}) \right\|_{L^2} \le \varepsilon \left\| f^{(r)} \right\|_{L^2}$$

gilt. Nach Voraussetzung ist der Transferoperator $P_{\varphi,\tilde{\varphi}} = \mathcal{E}_\varphi \mathcal{E}_{\tilde{\varphi}}^*$ beschränkt. Sei $B > 0$ eine gemeinsame obere Schranke der Prä–Gramschen Fasern J_φ und $J_{\tilde{\varphi}}$, dann ist B auch eine obere Schranke für $P_{\varphi,\tilde{\varphi}}$, und damit auch für jeden der skalierten Operatoren P_S. Mit der Dreiecksungleichung erhalten wir für die uns interessierende Differenz

$$S^r \left\| P_S(f) - f \right\|_{L^2} \le S^r (B+1) \left\| f - \pi_{\delta S}(f) \right\|_{L^2} + S^r \left\| P_S(\pi_{\delta S}(f)) - \pi_{\delta S}(f) \right\|_{L^2} .$$

Den ersten Summanden können wir wieder mit Hilfe der Plancherel–Identität gegen die r–te Ableitung abschätzen,

$$\begin{aligned} S^{2r} \left\| f - \pi_{\delta S}(f) \right\|_{L^2}^2 &= S^{2r} \int_{|\omega|>\delta S} |\hat{f}(\omega)|^2 \,\mathrm{d}\omega \le (2\pi\delta)^{-2r} \int_{|\omega|>\delta S} \left| (i2\pi\omega)^r \hat{f}(\omega) \right|^2 \mathrm{d}\omega \\ &= (2\pi\delta)^{-2r} \left\| f^{(r)} - \pi_{\delta S}(f^{(r)}) \right\|_{L^2}^2 . \end{aligned}$$

Zusammenfassend erhalten wir die Abschätzung

$$S^r \left\| P_S(f) - f \right\|_{L^2} \leq (1+B)(2\pi\delta)^{-r} \left\| f^{(r)} - \pi_{\delta S}(f^{(r)}) \right\|_{L^2} + \varepsilon \left\| f^{(r)} \right\|_{L^2} .$$

Bei $S \to \infty$ konvergiert der erste der Summanden auf der rechten Seite gegen Null, da $f^{(r)} \in L^2(\mathbb{R})$ vorausgesetzt war. Da aber $\varepsilon > 0$ beliebig war, kann auch der zweite Summand beliebig klein gehalten werden. Somit konvergiert die linke Seite gegen Null. □

Kapitel 5

Multiskalenanalyse

Um bandbeschränkte Prozesse zu analysieren bzw. beliebige Prozesse auf die Modellannahme der Frequenzbeschränktheit mit niedriger höchster Frequenz zu prüfen, könnte man versuchen, eine Messapparatur zu konstruieren, welche einen auf dem Kardinalsinus basierenden Abtastoperator realisiert. Eine gegenüber der zu testenden Grenzfrequenz genügend hohe Abtasttaktrate vorausgesetzt, kann die Messreihe dann gemäß einer Oktavbandzerlegung der Zeit–Frequenzebene (s. Abschnitt 5.1) in die Koeffizienten der Reihenentwicklung der Frequenzbänder umgerechnet werden. An diesen kann dann abgelesen werden, welche Frequenzbänder einen Beitrag zum gemessenen Prozess leisten, bzw. ob der Anteil über der Grenzfrequenz klein genug ist, um als Störung vernachlässigt zu werden.

Es gibt einige Probleme bei einem solchen Vorgehen. Der Kardinalsinus hat einen unbeschränkten Träger, womit eine praktische Realisierung dieser Funktion als Messprozess unmöglich ist. Dieses Problem kann durch Wechsel der Funktion im Abtastmodell und Anpassung der Abtastrate reduziert werden. Wählen wir beispielsweise die technisch einfach zu realisierende Blockfunktion χ_I, $I := [0, 1]$ in der Abtastung und behalten den Kardinalsinus in der Rekonstruktion bei, so wird die Kombination von Abtastung und Rekonstruktion im Allgemeinen nicht fehlerfrei sein. Der Fehler kann aber mittels *Überabtastung* (s. Satz 4.2.9), d.h. mittels Bestimmung der Zeitreihe mit wesentlich verkleinertem Zeitschritt, umgangen werden. Im Beispiel würde ein gegenüber dem idealen Modell um den Faktor 8 reduzierter Zeitschritt zu einem relativen Fehler kleiner 1% führen.

An diesem Punkt erwartet uns das nächste Problem, nämlich dass die Auswertung der mittels dieser Abtastung gewonnenen Folge die Bestimmung unendlich vieler Reihen erfordert. Auch hierfür kann eine in endlicher Zeit bestimmbare Näherung gefunden werden. Diese wird, da die in den Reihen vorkommenden Koeffizienten jedoch nur linear nach Unendlich abfallen, immer noch sehr aufwendig sein. Nachfolgend stellt sich dann die Frage einer frequenzbeschänkten Rekonstruktion…

An diesem Punkt angekommen, kann es sinnvoll erscheinen, den Begriff des Frequenzbandes in der Oktavbandzerlegung etwas „aufzuweichen“, um die Endlichkeit der Berechnungen und bessere, praktisch einfacher zu realisierende Funktionen für Abtastung und Rekonstruktion zu gewinnen. Der theoretische Rahmen für diese Konstruktion wird durch die *Multiska-*

lenanalyse gelegt.

Wir werden mit den *Haar–Wavelets* zunächst ein weiteres Beispiel einer Multiskalenanalyse betrachten, das in gewisser Weise den Gegenpol zur Oktavbandzerlegung bildet. Um die Multiskalenanalyse in der benötigten Allgemeinheit definieren zu können, benötigen wir noch einige funktionalanalytische Grundbegriffe. Aus den Forderungen an eine Multiskalenanalyse und deren Anwendung zur Konstruktion von Wavelet–Systemen ergeben sich analytische sowie algebraische Anforderungen an die Koeffizientenfolge der *Verfeinerungsgleichung*.

5.1 Zerlegungen der Zeit–Frequenz–Ebene

Unter einer *Zerlegung der Zeit–Frequenz–Ebene* wollen wir eine orthogonale Zerlegung des $L^2(\mathbb{R})$ in Unterräume $PW(A_n)$ bandbeschränkter Funktionen mit Frequenzbändern mit $A_n \subset \mathbb{R}$, $n \in \mathbb{N}$ bezeichnet ($\mathbb{N}$ kann durch andere abzählbare Indexmengen wie $\mathbb{Z}$ ersetzt werden). Jedes A_n sei abgeschlossen und beschränkt, der Schnitt $A_n \cap A_m$ zweier beliebiger Teilmengen sei eine Menge vom Maß Null. Der Raum $PW(A_n)$ besteht aus denjenigen Funktionen $f \in L^2(\mathbb{R})$, deren Fourier–Transformierte $\hat{f} := \mathcal{F}(f)$ ihren Träger in $\bar{A}_n$ hat.

Die Zerlegung $L^2(\mathbb{R}) = \bigoplus_{n\in\mathbb{N}} PW(A_n)$ ist genau dann eine orthogonale Zerlegung, wenn die Teilmengen A_n eine Zerlegung von $\mathbb{R}$ definieren, d.h. $\mathbb{R} = \bigcup_{n\in\mathbb{N}} A_n$ gilt, wobei die A_n bis auf Mengen vom Maß Null paarweise disjunkt sind. Damit kann jede Funktion $f \in L^2(\mathbb{R})$ auf die Unterräume $PW(A_n)$ projiziert werden. Bezeichnen wir die Projektionen mit $f_n \in PW(A_n)$, so gilt $f = \sum_{n\in\mathbb{N}} f_n$. Wird zu jedem A_n eine Bandbreite $B_n > 0$ gewählt, so kann f aus den Werten $f_n(\frac{m}{B_n})$, $(m,n) \in \mathbb{Z}\times\mathbb{N}$, rekonstruiert werden.

Eine allgemeine Klasse von Zerlegungen von $L^2(\mathbb{R},\mathbb{C})$ wird durch streng monotone, beidseitig unbeschränkte Folgen $a = \{a_n\}_{n\in\mathbb{Z}} \subset \mathbb{R}$ definiert. Dabei sind die Frequenzbänder durch die ganzen Zahlen indiziert und als $A_n := [a_n, b_n]$, $b_n := a_{n+1}$, definiert. Offensichtlich wird $\mathbb{R}$ durch diese Intervalle zerlegt.

Die Teilräume $PW([a_n,b_n]) \subset L^2(\mathbb{R})$ stehen nach der Plancherel–Identität und der Definition der Intervalle senkrecht zueinander. Mit der Bandbreite $B_n := (b_n - a_n)$ hat jedes $PW([a_n,b_n])$, $n \in \mathbb{Z}$, die Hilbert–Basis, die aus den Funktionen

$$\varphi_{n,m} := \frac{1}{B_n}\mathcal{T}_{\frac{m}{B_n}}\mathcal{F}^{-1}(\chi_{[a_n,b_n]}) = \mathcal{T}_{\frac{m}{b_n-a_n}}\left(e_{\frac{a_n+b_n}{2}}\mathcal{D}_{b_n-a_n}\,\mathrm{sinc}\right),$$

$m \in \mathbb{Z}$, besteht. Die Charakterisierung eines $f \in L^2(\mathbb{R})$ durch die Funktionswerte $\{f_n(\frac{m}{B_n}) : (m,n) \in \mathbb{Z}^2\}$ der Projektionen f_n von f auf die Teilräume ist dann eineindeutig und es gilt

$$f = \sum_{m,n\in\mathbb{Z}} f_n\left(\tfrac{m}{b_n-a_n}\right)\ \varphi_{n,m}$$

Die erzeugenden Funktionen $\varphi_{n,m} \in L^2(\mathbb{R})$, $m \in \mathbb{Z}$, bilden bei konstantem n eine Hilbert–Basis von $PW([a_n,b_n])$; das gesamte System $\{L_{n,m} : \ m,n \in \mathbb{Z}\}$ ist eine Hilbert–Basis des $L^2(\mathbb{R})$.

Jedem Indexpaar $(m,n) \in \mathbb{Z}^2$ sei in der Zeit–Frequenz–Ebene der Punkt $(\frac{m}{b_n-a_n}, \frac{a_n+b_n}{2}) \in \mathbb{R}^2$ zugeordnet. Dieser liegt in der Mitte des Rechtecks $\left[\frac{2m-1}{2(b_n-a_n)}, \frac{2m+1}{2(b_n-a_n)}\right] \times [a_n,b_n]$, alle der-

art definierten Rechtecke zusammen bilden eine Zerlegung des $\mathbb{R}^2$. Diese Zerlegung des $\mathbb{R}^2$ als Zeit–Frequenz–Ebene in Rechtecke kann als symbolische Darstellung der Hilbert–Basis $\{L_{n,m} : (m,n) \in \mathbb{Z}^2\}$ verstanden werden.

5.1.1 Reelle Zerlegungen der Zeit–Frequenz–Ebene

Unter einer reellen Zerlegung der Zeit–Frequenz–Ebene verstehen wir die Zerlegung des Raums $L^2(\mathbb{R},\mathbb{R})$ reellwertiger quadratintegrierbarer Funktionen in orthogonale Teilräume reellwertiger bandbeschränkter Funktionen $PW_{\mathbb{R}}(A_n)$. Da die Fourier–Transformierten reellwertiger Funktionen Träger haben, welche symmetrisch zum Nullpunkt liegen, ist dieses auch von den Intervallen A_n, $n \in \mathbb{N}$, zu fordern.

Im einfachsten Fall ist dies durch eine streng monoton wachsende, nach oben unbeschränkte Folge $\{a_n\}_{n\in\mathbb{N}}$ mit $a_0 = 0$ zu erreichen. Die Frequenzbänder sind dabei als $A_n = [-b_n, -a_n] \cup [a_n, b_n]$, $b_n := a_{n+1}$, für jedes $n \in \mathbb{N}$ definiert.

Nach Satz 4.1.9 haben die Räume $PW_{\mathbb{R}}(A_n)$ nur dann eine erzeugende Funktion, welche eine Hilbert–Basis dieses Teilraums erzeugt, wenn a_n ein ganzes Vielfaches von $(b_n - a_n) = (a_{n+1} - a_n)$ ist. Soll dies für alle $n \in \mathbb{N}$ gelten, muss es also eine Folge $\kappa := \{\kappa_n\}_{n\in\mathbb{N}_{>0}} \subset \mathbb{N}_{>0}$ geben, so dass $a_{n+1} = (1 + \frac{1}{\kappa_n})a_n$ für alle $n \in \mathbb{N}_{>0}$ gilt.

Eine einfache Variante, eine solche Folge zu konstruieren, ist durch die Intervallgrenzen $a_n := n\Omega$, $n \in \mathbb{N}$, mit einem fest vorgegeben $\Omega > 0$ gegeben. Dabei gilt $\kappa_n = n$, zu dieser Zerlegung gibt es eine Hilbert–Basis von $L^2(\mathbb{R},\mathbb{R})$. Die zugehörigen Frequenzbänder haben die gemeinsame Bandbreite $B = 2\Omega$, die erzeugende Funktion von $PW_{\mathbb{R}}(A_n)$ ist (s. Satz 4.1.9)

$$\begin{aligned}\varphi_n(x) =& (n+1)B\ \mathrm{sinc}((n+1)Bx) - NB\ \mathrm{sinc}(nBx) && \text{(5.1a)}\\ =& 2\Omega\ \cos(\pi(2n+1)\Omega x)\ \mathrm{sinc}(\Omega x)\,, && \text{(5.1b)}\end{aligned}$$

und jedes $f \in L^2(\mathbb{R},\mathbb{R})$ kann in eine Fourier–Reihe zur Hilbert–Basis $\{\mathcal{T}_{\frac{m}{B}}\varphi_n : (m,n) \in \mathbb{Z}^2\}$ entwickelt werden,

$$f = \sum_{(m,n)\in\mathbb{Z}^2} \left\langle f, \mathcal{T}_{\frac{m}{B}}\varphi_n \right\rangle \mathcal{T}_{\frac{m}{B}}\varphi_n\,.$$

Bemerkung: Diese Zerlegung der Frequenzachse entspricht einer idealisierten gefensterten Fourier–Transformation. Dabei ist durch $2\Omega\,\mathrm{sinc}(\Omega x)$ eine Fensterfunktion definiert, mit deren Verschiebungen um Vielfache von $\frac{1}{2\Omega}$ die reinen Schwingungen $\cos(\pi(2n+1)\Omega x)$, $n \in \mathbb{N}$, abgeschnitten werden.

Es hat sich als praktisch sinnvoll herausgestellt, auch Zerlegungen der Frequenzachse zu betrachten, in denen das Verhältnis von Frequenz und Bandbreite in den Frequenzbändern, d.h. die Glieder der Folge κ, beschränkt sind. Höheren Frequenzen soll also eine höhere Bandbreite B und damit eine geringere Schrittweite $\frac{1}{B}$ zugeordnet werden. Nach Grossmann/Morlet (s. [Dau96]) hat dies zum Hintergrund, dass z.B. in der Erforschung von Schwingungen bei Erdbeben hochfrequente Ereignisse eine kurze Zeitdauer haben, niederfrequente Schwingungen jedoch lange anhalten.

Im nächsten Abschnitt wird eine an diese Anforderungen angepasste Zerlegung der Zeit-Frequenz-Ebene, die Oktavbandzerlegung, konstruiert. Diese ist ein Modell für die Multiskalenanalyse.

5.1.2 Oktavbandzerlegung

Seien $\Omega > 0$ und $s \in \mathbb{N}_{>1}$ fest gewählt. Dann können wir die Forderung nach Beschränktheit des Verhältnisses von Frequenz zu Bandbreite durch eine periodische Folge κ mit $\kappa_{m+(s-1)n} := m$, $m = 1, \ldots, s-1$, $n \in \mathbb{N}$ erfüllen. Mit $a_0 = 0$ und $a_1 = \Omega$ erhalten wir dann die allgemeine Form eines Glieds der Folge $a = \{a_n\}_{n\in\mathbb{N}}$ als $a_{m+(s-1)n} = ms^n\Omega$, $m = 1, \ldots, s-1$, $n \in \mathbb{N}$.

Zur Aufzählung der Frequenzbänder verwenden wir der Übersicht halber statt der Menge $\mathbb{N}$ die Indexpaare der Menge $\{(0,0)\} \cup \{1, \ldots, s-1\} \times \mathbb{N}_{>0}$. Wir stellen jedes Frequenzband $A_{n,m}$ als Vereinigung zweier spiegelsymmetrischer Intervalle $(-I_{n,m}) \cup I_{n,m}$ dar, dabei ist

$$\begin{aligned} I_{0,0} :=& [0, \Omega], \\ I_{n,m} :=& [ms^n\Omega, (m+1)s^n\Omega], \qquad m = 1, \ldots, s-1, \qquad n \in \mathbb{N}_{>0}\,. \end{aligned}$$

Durch Variation von a_1 in der Menge $\{s^J\Omega : J \in \mathbb{Z}\}$ unter Beibehaltung der übrigen Bildungsvorschrift können wir nun den Anfang der Folge a modifizieren, während der Rest der Folge bis auf eine Indexverschiebung erhalten bleibt. Um diese Indexverschiebung in der Konstruktion der Intervalle zu vermeiden, schließen wir negative Indizes ein und definieren

$$\begin{aligned} I_{n,0} :=& [0, s^n\Omega], \\ I_{n,m} :=& [ms^n\Omega, (m+1)s^n\Omega], \qquad m = 1, \ldots, s-1, \qquad n \in \mathbb{Z}\,. \end{aligned}$$

Für jedes $J \in \mathbb{Z}$ und die Konstruktion der Folge a mit $a_1 = s^J\Omega$ erhalten wir eine Zerlegung des positiven Halbstrahls $\mathbb{R}_+$ durch

$$\mathbb{R}_+ = I_{J,0} \cup \bigcup_{\substack{n=J,J+1,\ldots \\ m=1,\ldots,s-1}} I_{n,m}$$

bzw. von ganz $\mathbb{R}$ unter Einschluss der jeweils gespiegelten Intervalle als

$$\mathbb{R} = A_{J,0} \cup \bigcup_{\substack{n=J,J+1,\ldots \\ m=1,\ldots,s-1}} A_{n,m}\,.$$

Im Fall $s = 2$ nennt man diese Zerlegung *Oktavbandzerlegung*, da die definierenden Intervalle $I_{j,1} = [2^j\Omega, 2^{j+1}\Omega]$, $j \in \mathbb{Z}$, eine Frequenzverdoppelung, d.h. eine Oktave im musikalischen Sinne repräsentieren.

Mit $J, K \in \mathbb{Z}$, $J < K$ gilt gleichfalls

$$I_{K,0} = I_{J,0} \cup (I_{J,1} \cup \cdots \cup I_{J,s-1}) \cup \cdots \cup (I_{K-1,1} \cup \cdots \cup I_{K-1,s-1})$$

und außerdem

$$\mathbb{R} \setminus \{0\} = \bigcup_{\substack{n\in\mathbb{Z} \\ m=1,\ldots,s-1}} A_{n,m}\,.$$

Bezeichnen wir die Räume bandbeschränkter Funktionen zu diesen Frequenzbändern mit

$$V_j := PW_{\mathbb{R}}\big(A_{j,0}\big), \quad W_{j,m} := PW_{\mathbb{R}}\big(A_{j,m}\big), \qquad m = 1, \dots, s-1, \quad j \in \mathbb{Z}\,,$$

so können wir aus den Zerlegungen folgende Eigenschaften der Teilräume von $L^2(\mathbb{R}, \mathbb{R})$ ableiten:

- Die Folge der Teilräume $\{V_j\}_{j\in\mathbb{Z}}$ ist aufsteigend, d.h. $V_j \subset V_{j+1}$.
- Die Teilräume V_j sind skalierte Versionen voneinander, d.h. $V_{j+1} = \mathcal{D}_s V_j$ für jedes $j \in \mathbb{Z}$.
- Der Teilraum V_0 wird durch ein Orthonormalsystem aufgespannt, welches unter Verschiebung mit Vielfachen von $\frac{1}{2\Omega}$ invariant ist.
- Die Vereinigung der Teilräume $\bigcup_{j\in\mathbb{Z}} V_j$ ist dicht in $L^2(\mathbb{R})$.
- Der Durchschnitt der Teilräume $\bigcap_{j\in\mathbb{Z}} V_j$ enthält nur die Nullfunktion.
- Das orthogonale Komplement von V_j in V_{j+1} ist durch $W_{j,1} \oplus \cdots \oplus W_{j,s-1}$ gegeben, d.h.

$$V_{j+1} = V_j \oplus W_{j,1} \oplus \cdots \oplus W_{j,s-1}$$

Dies sind die Eigenschaften, welche eine *orthogonale Multiskalenanalyse* des $L^2(\mathbb{R})$ durch eine aufsteigende Folge von Unterräumen definieren. Aus diesen folgt, dass für beliebige $J, K \in \mathbb{Z}$ mit $J < K$ gilt

$$V_K = V_J \oplus \bigoplus_{j=J}^{K-1} \big(W_{j,1} \oplus \cdots \oplus W_{j,s-1}\big)\ .$$

Im Grenzfall $J \to -\infty$, $K \to \infty$ erhalten wir in diesem Fall daraus eine orthogonale Zerlegung

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} \big(W_{j,1} \oplus \cdots \oplus W_{j,s-1}\big)\ .$$

Die Hilbert–Basen der Teilräume nach Satz 4.1.9 ergeben zusammengenommen eine Hilbert–Basis des $L^2(\mathbb{R})$. Diese entsteht aus den $s-1$ erzeugenden Funktionen der Räume $W_{0,1}, \dots, W_{0,s-1}$ durch Verschiebung um ganzzahlige Vielfache von $\frac{1}{2\Omega}$ und nachfolgende Dilatation mit allen ganzzahligen Potenzen von s. Eine solcherart strukturierte orthonormale Basis wird *Wavelet–Basis* genannt.

5.1.3 Analyse– und Syntheseoperatoren

Sei der Übersichtlichkeit halber hier nur der Fall $s = 2$ und $\Omega = \frac{1}{2}$ betrachtet. Dann haben, für jedes $j \in \mathbb{Z}$, die Unterräume V_j die Hilbert–Basen $\{2^{\frac{j}{2}} \mathcal{T}_k \mathcal{D}_2^j \varphi :\ k \in \mathbb{Z}\}$ mit $\varphi = \operatorname{sinc}$. Die komplementären Räume $W_j := W_{j,1}$ haben analog dazu die Hilbert–Basen $\{2^{\frac{j}{2}} \mathcal{T}_k \mathcal{D}_2^j \psi :\ k \in \mathbb{Z}\}$ mit $\psi = 2\mathcal{D}_2 \operatorname{sinc} - \operatorname{sinc}$. Nach Konstruktion sind sowohl φ als auch ψ in V_1 enthalten und lassen sich in der Basis von V_1, d.h. in eine Kardinalreihe entwickeln. Es gelten

$$\varphi(x) = \sum_{n\in\mathbb{Z}} \varphi(\tfrac{n}{2}) \operatorname{sinc}(2x-n) = \operatorname{sinc}(2x) + \sum_{m\in\mathbb{Z}} \operatorname{sinc}(m+\tfrac{1}{2}) \operatorname{sinc}(2x-2m-1)$$
$$\psi(x) = \sum_{n\in\mathbb{Z}} \psi(\tfrac{n}{2}) \operatorname{sinc}(2x-n) = \operatorname{sinc}(2x) - \sum_{m\in\mathbb{Z}} \operatorname{sinc}(m+\tfrac{1}{2}) \operatorname{sinc}(2x-2m-1)\ .$$

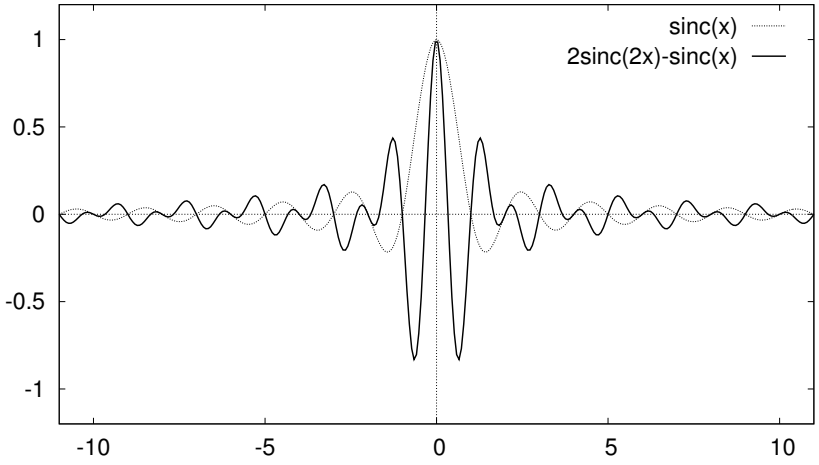


Abbildung 5.1: Die erzeugenden Funktionen der Hilbert–Basen von V_0 und W_0

Jede Funktion $f \in V_0$ kann sowohl in der Basis von V_0 als auch in der von V_1 ausgedrückt werden. So ist, im Einklang mit dem Additionstheorem des Kardinalsinus 4.1.7,

$$\begin{aligned} f(x) &= \sum_{n\in\mathbb{Z}} f(n)\,\mathrm{sinc}(x-n) = \sum_{k\in\mathbb{Z}} f\left(\tfrac{k}{2}\right)\mathrm{sinc}(2x-k) \\ &= \sum_{n\in\mathbb{Z}} f(n)\,\mathrm{sinc}(2x-2n) + \sum_{n\in\mathbb{Z}}\left(\sum_{m\in\mathbb{Z}} f(n-m)\,\mathrm{sinc}(m+\tfrac{1}{2})\right)\mathrm{sinc}(2x-2n-1)\,. \end{aligned}$$

Die Folge $c_1 := \{f(\frac{k}{2})\}_{k\in\mathbb{Z}}$ der Koeffizienten in der Basis von V_1 läßt sich also aus der Folge $c_0 := \{f(n)\}_{n\in\mathbb{Z}}$ der Koeffizienten der Basis von V_0 gewinnen. Eine Verschiebung von f um die Distanz Eins erzeugt in c_0 eine Verschiebung um ein Glied, in der Folge c_1 eine Verschiebung um zwei Glieder, die Abbildung von c_0 auf c_1 ist also $(2,1)$–periodisch. In der Polyphasenzerlegung von c_1 zum Faktor 2 ergibt sich die Teilfolge $\big(\downarrow 2\big)\, c_1$ direkt als c_0, für die zweite Teilfolge liest man ab

$$\big(\downarrow 2\big)(\mathcal{T}^{-1}c_1) = \{f(m+\tfrac{1}{2})\}_{m\in\mathbb{Z}} = \sum_{m\in\mathbb{Z}} \mathrm{sinc}(m+\tfrac{1}{2})\mathcal{T}^m c_0 = S_{\frac{1}{2}}c_0\,.$$

Dabei ist $S_{\frac{1}{2}}$ der in Korollar 4.1.7 definierte Operator

$$S_t := \sum_{m\in\mathbb{Z}} \mathrm{sinc}(m+t)\mathcal{T}^m : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C})$$

für $t = \frac{1}{2}$. $S_{\frac{1}{2}}$ ist unitär und es gilt $S_r S_t = S_{r+t}$ für beliebige $r, t \in \mathbb{R}$.

Fassen wir die Formeln der einzelnen Polyphasen zusammen, so gilt $\big(\downarrow\downarrow 2\big)\, c_1 = \Big(c_0, S_{\frac{1}{2}}c_0\Big)$. Genauso gibt es für jede Funktion $g \in W_0$ eine Entwicklung in der Hilbert–Basis zu W_0 mit Koeffizientenfolge $d_0 := \{g(n)\}_{n\in\mathbb{Z}}$ und eine Entwicklung in der Hilbert–Basis zu V_1 mit Koeffizientenfolge $d_1 = \{g(\frac{k}{2})\}_{k\in\mathbb{Z}}$. Analog zur vorhergehenden Rechnung gilt $\big(\downarrow\downarrow 2\big)\, d_1 = \Big(d_0, -S_{\frac{1}{2}}d_0\Big)$.

Bilden wir nun die Summe $f+g$, so hat diese in V_1 eine Entwicklung mit Koeffizientenfolge

$$\begin{aligned} c_1 + d_1 =& F(c_0, d_0) := \left(\uparrow\uparrow 2\right) \left(c_0 + d_0, S_{\frac{1}{2}}(c_0 - d_0)\right) \\ &= \left(\uparrow\uparrow 2\right) \begin{pmatrix} S_0 & 0 \\ 0 & S_{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} c_0 \\ d_0 \end{pmatrix} \end{aligned}$$

Jeder der Faktoren in obiger Zerlegung von F ist invertierbar, es gibt somit auch den inversen Operator

$$F^{-1} := \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} S_0 & 0 \\ 0 & S_{-\frac{1}{2}} \end{pmatrix} \left(\downarrow\downarrow 2\right) : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C})^2 \,,$$

mit welchem einer jeden Koeffizientenfolge $c_1 \in \ell_2(\mathbb{C})$ einer Kardinalreihe in V_1 die Koeffizientenfolgen c_0, d_0 der Entwicklung in den Hilbert–Basen von V_0 und W_0 zuordnet.

Im Rahmen der Theorie der Multiskalenanalyse wird die Identität

$$\varphi = \mathcal{D}_2 \sum_{k\in\mathbb{Z}} a_n \mathcal{T}^n \varphi \,,$$

mit der Koeffizientenfolge $a = F(\delta^0, 0) = \left(\uparrow\uparrow 2\right)(\delta^0, S_{\frac{1}{2}}\delta^0)^t$ als *Verfeinerungsgleichung* bezeichnet. Diese ist eine Funktionalgleichung, welche den Wert der Funktion φ an einer Stelle $x \in \mathbb{R}$ durch die Wertefolge als Linearkombination der Glieder der Folge $\{\varphi(2x+n)\}_{n\in\mathbb{Z}}$ ausdrückt. Auf gleiche Weise gilt

$$\psi = \mathcal{D}_2 \sum_{k\in\mathbb{Z}} b_k \mathcal{T}^k \psi$$

mit der Koeffizientenfolge $b = F(0, \delta^0) = \left(\uparrow\uparrow 2\right)(\delta^0, -S_{\frac{1}{2}}\delta^0)^t$.

Der $(2,1)$–periodische lineare Operator $F : \ell_2(\mathbb{C}^2) \to \ell_2(\mathbb{C})$ wird als *Synthese–Filterbank* bezeichnet. Mittels dieses Operators werden die Unterräume V_0 und W_0 zu V_1 zusammengeführt.

Umgekehrt wird der $(1,2)$–periodische lineare Operator $F^{-1} : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C}^2)$ als *Analyse–Filterbank* bezeichnet. Mit diesem Operator werden Funktionen in V_1 in Bestandteile aus V_0 und W_0 aufgespalten.

Die Möglichkeit, die Koeffizienten der Basen von $V_0 \oplus W_0$ direkt in Koeffizienten der Basis von V_1 umrechnen zu können, und umgekehrt, bildet die Grundlage der *schnellen Wavelet–Transformation*.

5.2 Haar–Wavelets

Eine einfach zu realisierende, jedoch analytisch ungünstige Möglichkeit einer die Oktavbandzerlegung des $L^2(\mathbb{R})$ ersetzenden Konstruktion sind die Haar–Wavelets (1909, nach Alfred Haar). Dabei werden die Teilräume bandbeschränkter Funktionen mit höchsten Frequenzen $s^j\Omega$, $j \in \mathbb{Z}$, durch die Teilräume der Treppenfunktionen der Stufenlängen s^jT, $j \in \mathbb{Z}$, ersetzt. Dabei sind $s \in \mathbb{N}_{>1}$, $\Omega, T > 0$ Parameter der jeweiligen Zerlegung des $L^2(\mathbb{R})$.

5.2.1 Treppenfunktionen

Sei $I := [0,1)$ ein Einheitsintervall. Die Funktionenfamilie $\{\mathcal{T}^n\chi_I\}_{n\in\mathbb{Z}} = \{\chi_{[n,n+1)}\}_{n\in\mathbb{Z}}$ bildet ein Orthonormalsystem in $L^2(\mathbb{R})$ mit dem Einbettungsoperator $\mathcal{E} : \ell_2(\mathbb{Z}) \to L^2(\mathbb{R})$ des Koordinatenraums $\ell_2(\mathbb{Z})$ und dem Projektor $P : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ auf den erzeugten Unterraum,

$$P(f) := \mathcal{E}\mathcal{E}^*(f) = \sum_{n\in\mathbb{Z}} \left(\int_n^{n+1} f(x)\,\mathrm{d}x \right) \cdot \chi_{[n,n+1)} \,.$$

Der von diesem Orthonormalsystem erzeugte Unterraum $V_0 := P(L^2(\mathbb{R}))$ von $L^2(\mathbb{R})$ ist derjenige, der alle Treppenfunktionen mit Stufenlänge 1 enthält. Andererseits wissen wir, dass sich beliebige Funktionen des $L^2(\mathbb{R})$ durch Treppenfunktionen beliebig genau approximieren lassen, wenn die Stufenlänge nur klein genug ist (s. auch Satz 4.2.9).

Sei ein $s \in \mathbb{N}$, $s > 1$ fixiert. Dann können wir die gestauchten Projektoren $P_j := \mathcal{D}_s^j P \mathcal{D}_s^{-j}$, $j \in \mathbb{Z}$, definieren. Diese ordnen jeder Funktion ihre orthogonale Projektion auf den Unterraum $V_j := P_j(L^2(\mathbb{R}))$ der Treppenfunktionen mit Stufenlänge s^{-j} zu,

$$P_j(f)(x) := \mathcal{D}_s^k \left(\sum_{n\in\mathbb{Z}} \left\langle \mathcal{D}_s^{-k} f,\, \mathcal{T}^n\chi_I \right\rangle \mathcal{T}^n\chi_I \right)(x) := \sum_{n\in\mathbb{Z}} \int_n^{n+1} f(s^{-k}t)\,\mathrm{d}t \cdot \chi_{[n,n+1)}(s^k x) \,.$$

Ist s^{-j} eine zum Erreichen einer genügend genauen Approximation einer Funktion $f \in L^2(\mathbb{R})$ notwendige Stufenlänge, so ist $P_j(f)$ diejenige Treppenfunktion aus V_j, welche den geringstmöglichen Abstand zu f realisiert. Die approximierenden Treppenfunktionen $P_j(f)$ konvergieren somit für jedes $f \in L^2(\mathbb{R})$ gegen die Funktion f, in $L^2(\mathbb{R})$ gilt $\lim_{k\to\infty} P_k f = f$.

5.2.2 Aufsteigende Folge von Unterräumen

Wir können für jedes $s \in \mathbb{N}_{>1}$ das Einheitsintervall $I := [0,1]$ in s Intervalle der Länge $\frac{1}{s}$ aufteilen, $\chi_I(x) = \chi_I(sx) + \chi_I(sx-1) + \cdots + \chi_I(sx-s+1)$. Das bedeutet nichts weiter, als dass eine Treppenfunktion $f = \mathcal{E}(c) = \sum_{n\in\mathbb{Z}} c_n \mathcal{T}_n \chi_I \in V_0$ mit Stufenlänge 1 ebenfalls eine Treppenfunktion in V_1 mit Stufenlänge $\frac{1}{s}$ ist, und damit identisch zu ihrer besten Approximation in V_1,

$$f = P_1 f = \sum_{n\in\mathbb{Z}} \sum_{j=0}^{s-1} c_n \mathcal{D}_s \mathcal{T}^{sn+j} \chi_I = \sum_{k\in\mathbb{Z}} \tilde{c}_k \mathcal{D}_s \mathcal{T}^k \chi_I = \mathcal{D}_s \mathcal{E}(\tilde{c}) \,.$$

Die Koeffizientenfolge $\tilde{c} = \{\tilde{c}_k\}_{k\in\mathbb{Z}}$ ist also auf Abschnitten der Länge s konstant. Eine andere Sichtweise ist, dass die Polyphasenteilfolgen von $\tilde{c}$ zum Faktor s Kopien von c sind. Dies bedeutet in der Umkehrung, dass

$$\tilde{c} = A(c) := \left(\uparrow\!\!\uparrow \mathrm{s}\right)(c, c, \ldots, c)^t = \left(\uparrow\!\!\uparrow \mathrm{s}\right)\left((1,1,\ldots,1)^t c\right)$$

gilt. Die $(s,1)$–periodische lineare Abbildung $A : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C})$ hat somit den mit Einsen gefüllten Spaltenvektor der Länge s als Polyphasenmatrix. Man überzeugt sich leicht, dass $\sqrt{s^{-1}}A$ semi–unitär ist, somit kann der Spaltenvektor $B^1 := A$ durch Ergänzen der Polyphasenmatrix so zu einem $(s,1)$–periodischen Operator $B : \ell_2(\mathbb{C}^s) \to \ell_2(\mathbb{C})$ erweitert werden, dass $\sqrt{s^{-1}}B$ unitär ist.

Eine der möglichen Erweiterungen der Polyphasenmatrix wird durch die Abbildungsmatrix der *diskreten Fourier–Transformation* (DFT) der Dimension s geliefert (s. Definition (4.3) im Beispiel auf Seite 121). $B_{poly} = (b_{j,k})_{j,k=1,\dots,s}$ hat dann die Einträge

$$b_{j,k} = e^{i2\pi s^{-1}(j-1)(k-1)}, \text{ d.h. } B_{poly} = \begin{pmatrix} 1 & q & q^2 & \dots & q^{s-1} \\ 1 & q^2 & q^4 & \dots & q^{2(s-1)} \\ \vdots & \vdots & & & \vdots \\ 1 & q^{s-1} & q^{2(s-1)} & \dots & q^{(s-1)^2} \end{pmatrix}$$

mit der primitiven Einheitswurzel $q = e^{i2\pi s^{-1}}$. Der inverse Operator $B^{-1} = s^{-1}B^* : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C}^s)$ ordnet jeder Koeffizientenfolge $\tilde{c}$ einer Funktion in V_1 Folgen $c, d^2, \dots, d^s \in \ell_2(\mathbb{C})$ zu.

Kombinieren wir $B : \ell_2(\mathbb{C}^s) \to \ell_2(\mathbb{C})$ mit der gestauchten Basisabbildung $\mathcal{D}_s\mathcal{E} : \ell_2(\mathbb{C}) \to V_1 \subset L^2(\mathbb{R})$, so erhalten wir eine Abbildung $\mathcal{D}_s\mathcal{E}B : \ell_2(\mathbb{C}^s) \to V_1$, welche $(1,1)$–periodisch ist in dem Sinne, dass einer Verschiebung um ein Glied in der Koeffizientenfolge eine Verschiebung um die Länge Eins in der Bildfunktion entspricht. Diese Abbildung ist also vollständig durch ihre Bildfolgen zu einelementigen Folgen bestimmt, seien also

$$\psi_k := \mathcal{D}_s\mathcal{E}B(\mathbf{e}_k\delta^0) = \mathcal{D}_s \sum_{n=1}^{s} b_{n,k}\mathcal{T}^{n-1}\varphi = \sum_{n=0}^{s-1} e^{i2\pi s^{-1}(k-1)n}\mathcal{D}_s\mathcal{T}^{n-1}\varphi\,.$$

Es gelten $\varphi = \psi_1$ und mit $(c, d^2, \dots, d^s)^t := B^{-1}(\tilde{c})$

$$\mathcal{D}_s\mathcal{E}(\tilde{c}) = \mathcal{E}(c) + \sum_{k=2}^{s}\sum_{n\in\mathbb{Z}} d_n^k\,\mathcal{T}^n\psi_k\,.$$

Die Funktionen $\psi_2, \dots \psi_s$ werden *Haar–Wavelets* genannt, (Haar untersuchte den Fall $s = 2$ vgl. [Dau92]). Das System $\{\sqrt{s^{-1}}\mathcal{T}^n\psi_k : k = 1, \dots, s,\ n \in \mathbb{Z}\}$ ist ebenfalls eine Hilbert–Basis von V_1, der von den Waveletfunktionen aufgespannte Unterraum

$$W_0 := \operatorname{span}\{\mathcal{T}^n\psi_k : k = 2, \dots, s,\ n \in \mathbb{Z}\} \subset V_1$$

ist das orthogonale Komplement zu V_0 in V_1.

Für jede Funktion $f \in L^2(\mathbb{R})$ kann also die Differenz zwischen den zwei aufeinander folgenden Projektionen P_0 und P_1 auf V_0 bzw. V_1 mittels dieser Haar–Wavelets ausgedrückt werden:

$$\begin{aligned} P_1(f) - P_0(f) &= \sum_{n\in\mathbb{Z}}\sum_{k=1}^{s} \tfrac{1}{s}\,\langle\, f,\, \mathcal{T}_n\psi_k\,\rangle\,\mathcal{T}_n\psi_k - \sum_{n\in\mathbb{Z}}\langle\, f,\, \mathcal{T}_n\varphi\,\rangle\,\mathcal{T}_n\varphi \\ &= \sum_{k=2}^{s}\sum_{n\in\mathbb{Z}} \tfrac{1}{s}\,\langle\, f,\, \mathcal{T}_n\psi_k\,\rangle\,\mathcal{T}_n\psi_k\,. \end{aligned} \tag{5.2}$$

5.2.3 Multiskalenanalyse

Bezeichnen wir mit W_j, $j \in \mathbb{Z}$, den durch Streckung oder Stauchung aus W_0 erhaltenen Unterraum

$$W_j := \mathcal{D}_s^j W_0 = \operatorname{span}\{\mathcal{D}_s^j\mathcal{T}^n\psi_k : k = 2, \dots, s,\ n \in \mathbb{Z}\}\,,$$

so gilt, analog zur Oktavbandzerlegung des $L^2(\mathbb{R})$, dass W_j das orthogonale Komplement zu V_j in V_{j+1} ist, d.h. $V_{j+1} = V_j \oplus W_j$. Diese Zerlegung der Unterräume läßt sich rekursiv fortsetzen, für beliebige $J, K \in \mathbb{Z}$ mit $K > J$ gilt

$$V_K = V_J \oplus \bigoplus_{j=J}^{K-1} \left(W_{j,1} \oplus \cdots \oplus W_{j,s-1} \right) .$$

Wir erhalten also wieder die Eigenschaften der Folge $\{V_j\}_{j\in\mathbb{Z}}$ von Unterräumen des $L^2(\mathbb{R})$, die wir schon bei der Oktavbandzerlegung als für eine Multiskalenanalyse typisch gekennzeichnet haben:

- Die Folge der Unterräume ist aufsteigend, $V_j \subset V_{j+1}$.
- Die Unterräume stehen durch Dilatation miteinander in Beziehung, $V_{j+1} = \mathcal{D}_s V_j$, sind also Versionen desselben Raumes auf verschiedenen Skalen.
- V_0 besitzt eine Hilbert–Basis, welche unter ganzzahligen Verschiebungen invariant ist, insgesamt gilt $\mathcal{T}^m V_0 = V_0$.
- Es gibt ein verschiebungsinvariantes Komplement W_j zu V_j in V_{j+1}, d.h. $V_{j+1} = V_j \oplus W_j$.
- Die Vereinigung $\bigcup_{j\in\mathbb{Z}} V_j$ ist dicht in $L^2(\mathbb{R})$. Es läßt sich ebenfalls zeigen, dass der Durchschnitt $\bigcap_{j\in\mathbb{Z}} V_j$ nur das Nullelement enthält.

5.3 Multiskalenanalyse

Wir erinnern an die auf $L^2(\mathbb{R})$ definierten Verschiebungs– und Skalierungsoperatoren. Wir bezeichnen mit $\mathcal{T}_t : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ den Translationsoperator zur Verschiebungsweite $t \in \mathbb{R}$ in Richtung wachsender Argumente, d.h. $(\mathcal{T}_t f)(x) := f(x - t)$. Weiterhin bezeichnet $\mathcal{D}_s : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ für jedes $s \in \mathbb{R}_+$ den Dilatationsoperator, welcher als $(\mathcal{D}_s)f(x) := f(sx)$ definiert ist. Im Gegensatz zur Translation, welche isometrisch ist, erhält die Dilatation die Norm in $L^2(\mathbb{R})$ nur bis auf einen konstanten Faktor, es gilt $\|\mathcal{D}_s f\|_{L_2} = \frac{1}{\sqrt{s}} \|f\|_{L_2}$. Von beiden Operatoren können ganzzahlige Potenzen gebildet werden, es gelten $\mathcal{T}_t^m = \mathcal{T}_{mt}$ und $\mathcal{D}_s^m = \mathcal{D}_{s^m}$ für jedes $m \in \mathbb{Z}$. Desweiteren gilt für die Verknüpfung beider Operatoren $\mathcal{T}_t \mathcal{D}_s = \mathcal{D}_s \mathcal{T}_{st}$. Wir vereinbaren $\mathcal{T} = \mathcal{T}_1$, d.h. $\mathcal{T}^n = \mathcal{T}_n$.

Definition 5.3.1 *Eine* Multiskalenanalyse *des Raums $L^2(\mathbb{R})$ besteht aus einer Folge rekursiv eingebetteter Unterräume*

$$\{0\} \subset \cdots \subset V_0 \subset V_1 \subset \cdots \subset V_n \subset V_{n+1} \subset \cdots \subset L^2(\mathbb{R}) ,$$

welche folgende Bedingungen erfüllt:

- *Es gibt ein $s \in \mathbb{N}$, $s > 1$ mit $V_{k+m} = \mathcal{D}_s V_k$ für alle $k, m \in \mathbb{Z}$. Das heißt, zu jeder Funktion $f \in V_k$ gibt es eine Funktion $g \in V_{k+m}$ mit $g(x) = f(s^m x)$.*

- *Mit dem Faktor s gilt $V_k = \mathcal{T}_{s^{-k}m} V_k$ für jedes $k \in \mathbb{Z}$ und für jedes $m \in \mathbb{Z}$. Insbesondere gilt $V_0 = \mathcal{T}_m V_0$. Dies heißt, zu jeder Funktion $f \in V_k$ und zu jedem $m \in \mathbb{Z}$ gibt es eine Funktion $g \in V_k$ mit $f(x) = g(x + ms^{-k})$.*
- *Im Unterraum V_0 gibt es endlich viele Funktionen $\varphi_1, \ldots, \varphi_r$, $r \in \mathbb{N}$, so dass das verschiebungsinvariante System*

$$X_0 := X(\{\varphi_1, \ldots, \varphi_r\}) := \{\mathcal{T}^k \varphi_i : \ i = 1, \ldots, r, \ k \in \mathbb{Z}\}$$

  eine Riesz–Basis von $V_0 \subset L^2(\mathbb{R})$ ist.
- *Die Vereinigung $\bigcup_{k\in\mathbb{Z}} V_k$ aller Unterräume ist eine dichte Teilmenge des $L^2(\mathbb{R})$ und der Durchschnitt $\bigcap_{k\in\mathbb{Z}} V_k$ ist der Nullunterraum $\{0\}$.*

Der Faktor s wird Skalenfaktor und die Funktionen $\varphi_1, \ldots, \varphi_r$ werden *Skalierungsfunktionen* oder *Vaterwavelets* genannt.

Als Beispiele haben wir bereits das Haar–Wavelet–System und die bandbeschränkten Waveletsysteme kennengelernt. Im Haar–Waveletsystem ist der Raum V_0 durch die Treppenfunktionen gegeben, die jeweils auf allen ganzzahligen Intervallen $(n, n+1)$ konstant sind. Die Skalierungsfunktion $\varphi := \chi_{[0,1)}$ erzeugt dann eine Hilbert–Basis des Raums V_0.

Ebenfalls erzeugt der Kardinalsinus ein Orthonormalsystem, der von dessen Verschiebungen aufgespannte Unterraum $V_0 = PW([-\frac{1}{2}, \frac{1}{2}])$ ist der Raum der bandbeschränkten Funktionen mit höchster Frequenz $\frac{1}{2}$. Die Räume V_j enthalten die bandbeschränkten Funktionen mit höchster Frequenz 2^{j-1}, schöpfen also den Raum $L^2(\mathbb{R})$ aus.

Aus der Inklusion $V_0 \subset V_1$ ergibt sich, dass es Folgen $a_{i,k} \in \ell_2(\mathbb{Z})$, $i, k = 1, \ldots, r$ geben muss mit

$$\varphi_i = \sum_{k=1}^{r} \sum_{n\in\mathbb{Z}} a_{i,k,n} \mathcal{D}_s \mathcal{T}^n \varphi_k, \qquad i = 1, \ldots, r\,.$$

Die Folgen $a_{i,k} \in \ell_2(\mathbb{Z}, \mathbb{C})$ bzw. die aus ihnen zusammengesetzte matrixwertige Folge $a \in \ell_2(\mathbb{Z}, \mathbb{C}^{r\times r})$ werden als *Skalierungsfolge* bezeichnet.

Wir werden uns im folgenden auf den Fall der *einfachen* Wavelets, also auf den Fall $r = 1$, einschränken. Es ist üblich, nur solche Funktionen $\varphi \in L^2(\mathbb{R})$ als Skalierungsfunktionen anzusehen, deren Skalierungsfolge a in $\ell_1(\mathbb{Z})$ liegt. Somit ist der Differenzenoperator

$$a(\mathcal{T}) := \sum_{n\in\mathbb{Z}} a_n \mathcal{T}^n : L^2(\mathbb{R}) \to L^2(\mathbb{R})$$

beschränkt und φ erfüllt die *Verfeinerungsgleichung*

$$\varphi = \mathcal{D}_s a(\mathcal{T}) \varphi\,, \tag{5.3}$$

Punktweise gilt also $\varphi(x) = \sum_{n\in\mathbb{Z}} a_n \varphi(sx - n)$ für fast jedes $x \in \mathbb{R}$. Diese Einschränkung ist für die Haar–Wavelets, aber nicht für die bandbeschränkten Wavelets erfüllt, da der Kardinalsinus als für jedes $s \in \mathbb{N}_{>1}$ eine Verfeinerungsgleichung mit einer Skalierungsfolge

$$\left\{\operatorname{sinc}(\tfrac{k}{s})\right\}_{k\in\mathbb{Z}} = \left(\uparrow\!\!\uparrow \mathrm{s}\right) \left(S(0), S(\tfrac{1}{s}), \ldots, S(\tfrac{s-1}{s})\right)^t$$

erfüllt. Da die Polyphasenteilfolgen die harmonische Reihe als Minorante haben, ist diese Skalierungsfolge nicht in $\ell_1(\mathbb{C})$ enthalten und damit nicht zulässig.

5.3.1 Zulässige Skalierungsfunktionen

Die analytischen Eigenschaften einer Multiskalenanalyse lassen sich im Fall einer einzigen Skalierungsfunktion φ direkt auf die Eigenschaften des von φ erzeugten verschiebungsinvarianten Systems zurückführen. Die hier entscheidenden Eigenschaften sind, dass dieses ein Riesz–System ist und dass φ eine Approximationsbedingung nach Definition 4.2.5 erfüllt.

Lemma 5.3.2 (s. [Dau92], s.a. [Mic91]) *Seien $\varphi \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ und $s \in \mathbb{N}$, $s > 1$ gegeben. Sei das von φ erzeugte verschiebungsinvariante System $X_0 := X(\varphi)$ ein Riesz–System mit Schranken $0 < A \leq B < \infty$.*

Wir betrachten die Funktionenfamilien $X_j := \{\varphi_{j,k}\}_{k\in\mathbb{Z}}$, $j \in \mathbb{Z}$, deren Elemente definiert sind durch $\varphi_{j,k}(x) := s^{j/2}(\mathcal{D}_s^j \mathcal{T}^k \varphi)(x) = s^{j/2}\varphi(s^j x - k)$. Sei mit $V_j := \operatorname{span}(X_j)$ der von X_j aufgespannte Teilraum von $L^2(\mathbb{R})$ bezeichnet.

Dann ist für jedes $j \in \mathbb{Z}$ die Teilmenge $X_j \subset L^2(\mathbb{R})$ ein Riesz–System mit denselben Schranken $0 < A \leq B < \infty$ und der Durchschnitt der aufgespannten Teilunterräume ist trivial, es gilt $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

Beweis: Bezeichnen wir die Synthese–Operatoren der Systeme kurz mit $\mathcal{E}_j := \mathcal{E}_{X_j}$, so gilt $\mathcal{E}_j = s^{\frac{j}{2}} \mathcal{D}_s^j \mathcal{E}_0$ und damit für jedes $c \in \ell_2(\mathbb{Z})$

$$\|\mathcal{E}_j(c)\|_{L^2}^2 = \left\| s^{\frac{j}{2}} \mathcal{D}_s^j \mathcal{E}_0(c) \right\|_{L^2}^2 = \|\mathcal{E}_0(c)\|_{L^2}^2 \in \left[A\|c\|_{\ell_2}^2,\ B\|c\|_{\ell_2}^2 \right] .$$

Daher ergeben die Abschätzungen für X_0 auch identische Ungleichungen für X_j. Weiterhin ist jedes Riesz–System eine Riesz–Basis des von ihm aufgespannten Unterraums. Insbesondere gilt für jedes $f \in V_j$ auch die Frame–Bedingung

$$A\|f\|_{L^2}^2 \leq \|\mathcal{E}_j^*(f)\|_{\ell_2}^2 = \sum_{k\in\mathbb{Z}} |\left\langle f,\, \varphi_{j,k} \right\rangle|^2 \leq B\|f\|_{L^2}^2 .$$

Sei $f \in \bigcap_{j\in\mathbb{Z}} V_j$. Dann gibt es zu jedem $j \in \mathbb{Z}$ eine Folge $c_j \in \ell_2(\mathbb{Z})$ mit $f = \mathcal{E}_j(c_j) = \sum_{k\in\mathbb{Z}} c_{j,k}\varphi_{j,k}$. Sei $g \in C_c(\mathbb{R})$ eine beliebige stetige Funktion mit kompaktem Träger, d.h. es gebe ein $R > 0$ mit $\operatorname{supp}\varphi \subset [-R, R]$. Dann gilt

$$|\langle f, g\rangle_{L^2}| = \left| \left\langle c_j,\, \mathcal{E}_j^*(g) \right\rangle_{\ell_2} \right| \leq \|c_j\|_{\ell_2} \left\| \mathcal{E}_j^*(g) \right\|_{\ell_2} = \|c_j\|_{\ell_2} \sqrt{\sum_{k\in\mathbb{Z}} \left| \left\langle \varphi_{j,k},\, g \right\rangle_{L^2} \right|^2} .$$

Der erste Faktor kann nach den Schranken des Riesz–Systems abgeschätzt werden, $\|c_j\|_{\ell_2} \leq \frac{1}{\sqrt{A}}\|f\|$. Für die Skalarprodukte im zweiten Faktor gilt

$$|\left\langle \varphi_{j,k},\, g \right\rangle| = \left| \left\langle \chi_{[-R,R]}\varphi_{j,k},\, g \right\rangle \right| \leq \|\chi_{[-R,R]}\varphi_{j,k}\|_{L^2}\, \|g\|_{L^2} ,$$

so dass in der Abschätzung des Skalarprodukts $\langle f, g \rangle_{L^2}$ nur noch die Norm der Folge $\{\|\chi_{[-R,R]}\varphi_{j,k}\|_{L^2}\}_{k\in\mathbb{Z}}$ von V_j abhängt. Es gilt für alle $j,k \in \mathbb{Z}$ nach der Definition der Funktionenfamilie $\{\varphi_{j.k} : j,k \in \mathbb{Z}\}$

$$\|\chi_{[-R,R]}\varphi_{j,k}\|_{L^2} = \|\chi_{[-s^jR,s^jR]}\varphi_{0,k}\|_{L^2} = \|\chi_{[k-s^jR,k+s^jR]}\varphi\|_{L^2} .$$

Ist $J \in \mathbb{Z}$ klein genug, so gilt für jedes $j \leq J$ die Ungleichung $s^jR < \frac{1}{2}$. Dann sind die Intervalle $[k - s^2R, k + s^jR]$, $k \in \mathbb{Z}$, paarweise disjunkt, und die Funktionen des Systems $\{\chi_{[k-s^jR,k+s^jR]}\varphi : k \in \mathbb{Z}\}$ stehen paarweise senkrecht zueinander. Nach dem Satz des Pythagoras gilt somit

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \|\chi_{[-R,R]}\varphi_{j,k}\|_{L^2}^2 = \sum_{k\in\mathbb{Z}} \|\chi_{[k-s^jR,k+s^jR]}\varphi\|_{L^2}^2 &= \left\|\sum_{k\in\mathbb{Z}} \chi_{[k-s^jR,k+s^jR]}\varphi\right\|_{L^2}^2 \\ &= \left\|\chi_{D_j}\varphi\right\|_{L^2}^2 .\end{aligned}$$

Dabei bezeichneten wir die Vereinigung der disjunkten Intervalle als $D_j := \bigcup_{k\in\mathbb{Z}}[k - s^jR, k + s^jR]$. Diese Menge ist bei gegen $-\infty$ strebendem j auf stets kleinere Umgebungen der ganzen Zahlen beschränkt, d.h. wir schneiden zunehmend „größere Löcher“ in den Definitionsbereich der Funktion φ. Somit kann zusammenfassend geschrieben werden

$$|\langle f, g \rangle| \leq \frac{1}{\sqrt{A}}\|f\|_{L^2}\,\|g\|_{L^2}\,\|\chi_{D_j}\varphi\|_{L^2} .$$

Die Folgen $\{\chi_{D_{-J-n}}(x)\}_{n\in\mathbb{N}}$ nehmen für jedes $x \in \mathbb{R}$ nur die Werte 0 und 1 an, sind monoton fallend und werden für jedes nicht in $\mathbb{Z}$ enthaltene x bei 0 stationär. Daher ist nach dem Satz von Lebesgue über die dominierte Konvergenz die Folge $\{\chi_{D_{-j}}\varphi\}_{j\in\mathbb{N}}$ eine Nullfolge in $L^2(\mathbb{R})$. Somit ist $\langle f, g \rangle = 0$, und dies für jede stetige Funktion mit kompaktem Träger. Also muss schon $f = 0$ gelten. □

Definition 5.3.3 *Eine Funktion $\varphi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ heiße* zulässige Skalierungsfunktion, *wenn es einen Faktor $s \in \mathbb{N}_{>1}$ und eine Folge $a \in \ell_1(\mathbb{Z})$ derart gibt, dass*

i) die Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ erfüllt ist und

ii) φ eine Approximationsbedingung erfüllt, d.h. für die Prä–Gramsche Faser wenigstens $J_\varphi = \delta_0 + o_{\ell_2}(1)$ gilt.

Satz 5.3.4 *Sei φ eine zulässige Skalierungsfunktion zum Skalenfaktor $s \in \mathbb{N}_{>1}$, so dass das von φ erzeugte verschiebungsinvariante System $X_0 := X(\varphi)$ ein Riesz–System mit Schranken $0 < A \leq B < \infty$ ist. Dann erzeugt φ eine Multiskalenanalyse $\{V_j\}_{j\in\mathbb{Z}}$ mit $V_j = \mathcal{D}_s^j V_0$ und $V_0 = \operatorname{span}(X_0)$.*

Beweis: Nach Satz C.3.4 ist die Eigenschaft von φ, ein Riesz–System zu erzeugen, äquivalent zur beidseitigen Beschränktheit der Prä–Gramschen Fasern $J_\varphi(\omega) \in \ell_2(\mathbb{Z})$, d.h. für fast jedes $\omega \in \mathbb{R}$ gilt $A \leq \|J_\varphi(\omega)\|^2 \leq B$.

Mit der vorausgesetzten Approximationsbedingung kann nach Satz 4.2.9 jede Funktion in $L^2(\mathbb{R})$ beliebig gut durch die Unterräume V_j approximiert werden, d.h. $\bigcup_{j\in\mathbb{Z}} V_j$ ist dicht in

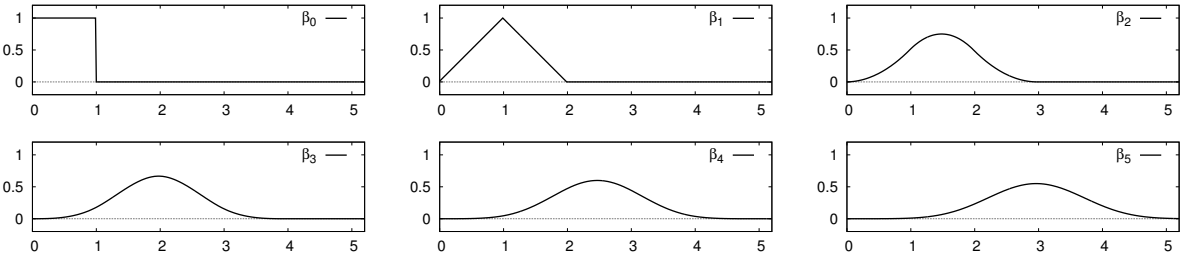


Abbildung 5.2: B–Splines der Ordnungen 0 bis 5

$L^2(\mathbb{R})$. Weiter folgt aus der Existenz der unteren Schranke $A > 0$ nach dem vorhergehenden Lemma 5.3.2 für den Durchschnitt der Unterräume $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

Seien wieder die Funktionenfamilien $X_j = \{\varphi_{j,k} : \ k \in \mathbb{Z}\}$ definiert, deren Elemente durch $\varphi_{j,k} = s^{j/2}\mathcal{D}_s^j\mathcal{T}^k\varphi$, d.h. $\varphi_{j,k}(x) = s^{j/2}\varphi(s^j x - k)$, gegeben sind. X_0 ist eine Riesz–Basis von $V_0 := \operatorname{span}(X_0)$, und damit auch X_j eine Riesz–Basis von $V_j = \operatorname{span}(X_j)$ für jedes $j \in \mathbb{Z}$.

Es verbleibt noch die Schachtelung der Folge der Unterräume $\{V_j\}_{j\in\mathbb{Z}}$ nachzuweisen. Dazu ist es ausreichend zu zeigen, dass $V_0 \subset V_1 = \operatorname{span}(X_1)$ ist. Sei dazu ein beliebiges $f \in V_0$ fixiert. Dann gibt es also eine Folge $c_0 \in \ell_2(\mathbb{Z})$ mit

$$f = \sum_{k\in\mathbb{Z}} c_{0,k}\varphi_{0,k} = \sum_{k\in\mathbb{Z}} c_{0,k}\mathcal{T}^k\mathcal{D}_s a(\mathcal{T})\varphi = \mathcal{D}_s\left(\sum_{k\in\mathbb{Z}} c_{0,k}\mathcal{T}^{sk}a(\mathcal{T})\right)\varphi\,.$$

f ist auch in V_1 enthalten, wenn die neue Koeffizientenfolge $c_1 := a(\mathcal{T})\,(\uparrow\mathrm{s})\,c_0$ quadratsummierbar ist. Das ist der Fall, da folgende Abschätzung gilt.

$$\|c_1\|_{\ell_2} \le \|a\|_{\ell_1}\|\,(\uparrow\mathrm{s})\,c_0\|_{\ell_2} = \|a\|_{\ell_1}\|c_0\|_{\ell_2}$$

□

5.3.2 B–Splines

Ein wichtiges Hilfsmittel der Lösungstheorie für die Verfeinerungsgleichung sowie an sich wichtige Beispiele für Skalierungsfunktionen und zugehörige Multiskalenanalysen sind die *B–Splines*.

Definition 5.3.5 *Die Familie* $\{\beta_m : m \in \mathbb{N}\}$ *der* B–Splines *ist rekursiv definiert durch* $\beta_0 := \chi_{[0,1)}$ *und* $\beta_{m+1} = \beta_m * \beta_0$ *für jedes* $m \in \mathbb{N}$; β_m *wird B–Spline der* Ordnung m *genannt.*

Man überzeugt sich leicht, dass aus der rekursiven Definition folgt, dass Faltungsprodukte von B–Splines wieder solche sind. Es gilt $\beta_m * \beta_n = \beta_{m+n+1}$ für alle $m, n \in \mathbb{N}$.

Lemma 5.3.6 *Es gilt mit* $x_+ = \max(x, 0)$ *die Darstellung*

$$\beta_m(x) = (1 - \mathcal{T})^{m+1}\frac{(x_+)^m}{m!}\,,$$

und mit $s \in \mathbb{N}_{>1}$ die Verfeinerungsgleichung

$$\beta_m = \mathcal{D}_s \left(s\, H_s(\mathcal{T})^{m+1} \beta_m \right) , \tag{5.4}$$

wobei $H_s(Z) := \frac{1}{s}\left(1 + Z + \cdots + Z^{s-1}\right)$ wieder das Haar–Polynom *ist. Die Fourier–Transformierte des B–Splines der Ordnung m ist*

$$\widehat{\beta_m}(\omega) = e_{m+1}(-\tfrac{\omega}{2})\, \mathrm{sinc}(\omega)^{m+1}$$

Beweis: $\beta_0 = \chi_{[0,1)}$ ist die Differenz zweier Sprungfunktionen mit Sprüngen bei 0 und 1, $\beta_0(x) = x_+^0 - (x-1)_+^0 = (1-\mathcal{T})(x_+)^0$. Für die Faltung von Potenzen von x_+ erhalten wir

$$\left((x_+)^m * (x_+)^0\right)(x) = \int_0^x t^m \,\mathrm{d}t = \frac{1}{m+1}(x_+)(x)^{m+1} ,$$

was sich durch Rekursion leicht erweitern lässt zu

$$\frac{(x_+)^m}{m!} * \frac{(x_+)^n}{n!} = \frac{(x_+)^{m+n+1}}{(m+n+1)!} .$$

Somit gilt nach Definition

$$\beta_m(x) = \underbrace{(1-\mathcal{T})x_+^0 * \cdots * (1-\mathcal{T})x_+^0}_{m+1 \text{ Faktoren}} = (1-\mathcal{T})^{m+1}\frac{(x_+)^m}{m!} .$$

Betrachten wir das Verhalten der B–Splines unter Dilatation. Es ist $\mathcal{D}_s(x_+) = (sx)_+ = s\, x_+$, da $s > 0$. Wir erhalten somit, unter der Bedingung $s \in \mathbb{N}$, $s > 1$,

$$\begin{aligned}\mathcal{D}_s^{-1}\beta_m =& \mathcal{D}_s^{-1}\left((1-\mathcal{T})^{m+1}\frac{(x_+)^m}{m!}\right) = (1-\mathcal{T}^s)^{m+1}\mathcal{D}_s^{-1}\left(\frac{(x_+)^m}{m!}\right) = (1-\mathcal{T}^s)^{m+1}\frac{(x_+)^m}{s^m\, m!} \\ =& s\left(\frac{1+\mathcal{T}+\cdots+\mathcal{T}^{s-1}}{s}\right)^{m+1}(1-\mathcal{T})^{m+1}\frac{(x_+)^m}{m!} = s\, H_s(\mathcal{T})^{m+1}\beta_m(x) .\end{aligned}$$

Für eine beliebige Funktion $f \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ ist $f * \chi_{[0,1)}$ wieder ein Element dieses Funktionenraums. Die Fouriertransformierte des Faltungsproduktes bestimmt sich als

$$\begin{aligned}\widehat{f * \chi_{[0,1)}}(\omega) =& \int_{\mathbb{R}}\int_0^1 f(x-t)\,\mathrm{d}t e^{-i2\pi\omega x}\,\mathrm{d}x = \int_0^1\int_{\mathbb{R}} f(x-t)e^{-i2\pi\omega(x-t)}\,\mathrm{d}x e^{-i2\pi\omega t}\,\mathrm{d}t \\ =& \hat{f}(\omega)\frac{e^{-i2\pi\omega}-1}{-i2\pi\omega} = e_1(-\tfrac{\omega}{2})\,\mathrm{sinc}(\omega)\hat{f}(\omega) .\end{aligned}$$

Die Behauptung zur Fourier–Transformierten folgt durch Induktion. □

Beispiel: Zwei einfache Verfeinerungsrelationen mit niedrigen Parametern sind

$$\begin{aligned}\beta_0(\tfrac{1}{3}x) =& 3H_3(\mathcal{T})\beta_0(x) = \beta_0(x) + \beta_0(x-1) + \beta_0(x-2) , \\ \beta_1(\tfrac{1}{2}x) =& 2H_2(\mathcal{T})^2\beta_0(x) = \tfrac{1}{2}\beta_1(x) + \beta_1(x-1) + \tfrac{1}{2}\beta_1(x-2) .\end{aligned}$$

Die zweite Relation ist in Abbildung 5.3 skizziert.

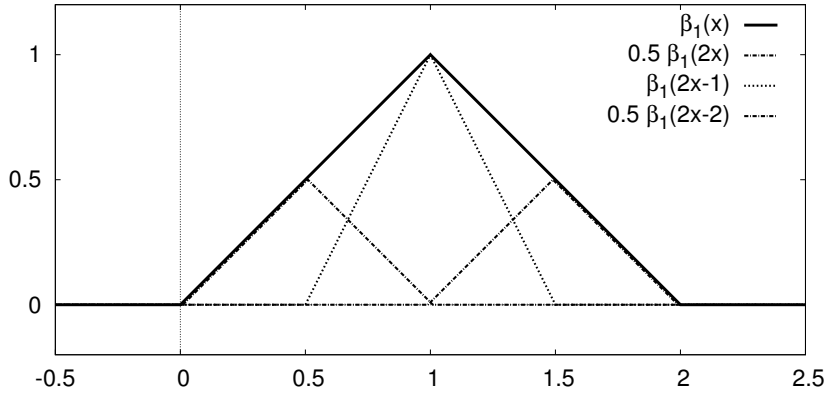


Abbildung 5.3: Verfeinerungsrelation der B–Splines am Beispiel des „Huts“ β_1

Satz 5.3.7 *Jeder B–Spline β_{A-1}, $A \in \mathbb{N}$, $A > 0$, erzeugt zu jedem Skalenfaktor $s \in \mathbb{N}$, $s > 1$, eine Multiskalenanalyse.*

Beweis: Sei die Ordnung $m \in \mathbb{N}$ des B–Splines fixiert. Wir prüfen die Voraussetzungen von Satz 5.3.4. Nach dem vorangehenden Lemma genügt β_{m-1} einer Verfeinerungsgleichung mit endlicher Skalierungsfolge. Alle B–Splines sind stückweise stetig mit kompaktem Träger, also in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ enthalten.

Es gilt für jedes $\omega \in [-\frac{1}{2}, \frac{1}{2}]$ und $n \in \mathbb{Z} \setminus \{0\}$ die Abschätzung

$$\operatorname{sinc}(\omega + n)^{2m+2} = \left(\frac{\sin(\pi\omega)}{\pi(\omega + n)}\right)^{2m+2} \leq \omega^{2m+2} \frac{1}{\left(|n| - \frac{1}{2}\right)^{2m+2}} \,. \tag{5.5}$$

Für die Prä–Gramsche Faser von β_m erhalten wir daraus

$$\left\| J_{\beta_m}(\omega) - e_{m+1}(-\tfrac{\omega}{2}) \operatorname{sinc}(\omega)^{m+1} \delta_0 \right\|_{\ell_2}^2 \leq 2(2\omega)^{2m+2} \sum_{n=1}^{\infty} \frac{1}{(2n-1)^{2m+2}} \,.$$

Da die darin vorkommende Reihe konvergiert, gilt $J_{\beta_m}(\omega) - \operatorname{sinc}(\omega)\delta_0 = O_{\ell_2}(\omega^{m+1})$. Der B–Spline der Ordnung m erfüllt somit eine Approximationsbedingung der Ordnung m (s. Definition 4.2.5). Insbesondere ist die Prä–Gramsche Faser nach oben beschränkt.

Die Norm der Prä–Gramschen Faser ist periodisch mit Periode 1, sie ist auf dem Intervall $[-\frac{1}{2}, \frac{1}{2}]$ durch $\operatorname{sinc}(\frac{1}{2})^{m+1}$ nach unten beschränkt, somit also auch überall.

Nach Satz 5.3.4 erzeugt also β_m eine Multiskalenanalyse. □

5.3.3 Weitere notwendige Bedingungen an Skalierungsfolge und –funktion

Wir können nach Lemma 4.2.4 zu jeder Ordnung $m \in \mathbb{N}$ eines B–Splines eine endliche Folge $g_m \in \ell_{\text{fin}}(\mathbb{C})$ finden, mit welcher für die Fourier–Transformierte von β_m die Entwicklung

$$\hat{g}_m \hat{\beta}_m = \hat{g}_m \left(e_1(-\tfrac{\omega}{2}) \operatorname{sinc}(\omega)\right)^{m+1} = 1 + o_{\mathbb{C}}(\omega^r)$$

gilt. Damit gilt aber auch für die Prä–Gramsche Faser $\hat{g}_m J_{\beta_m} - \delta_0 = o_{\ell_2}(\omega^r)$. Für Funktionen, welche eine Approximationsbedingung der Ordnung r nach Definition 4.2.5 erfüllen, erhalten wir somit folgende Charakterisierung.

Satz 5.3.8 *$\varphi \in L^2(\mathbb{C})$ erfüllt genau dann eine Approximationsbedingung der Ordnung $r \in \mathbb{N}$, wenn es für jedes $m \in \mathbb{N}$ mit $m \geq r$ eine endliche Folge $c \in \ell_{\text{fin}}(\mathbb{C})$ gibt, so dass*

$$J_\varphi(\omega) - \hat{c}(\omega) J_{\beta_m}(\omega) = o_{\ell_2}(\omega^r)$$

gilt.

Beweis: Erfüllt $\varphi \in L^2(\mathbb{C})$ eine Approximationsbedingung der Ordnung r, so gibt es nach Definition eine endliche Folge $c_0 \in \ell_{\text{fin}}(\mathbb{C})$ mit $J_\varphi = \hat{c}_0 \delta_0 + o_{\ell_2}(\omega^r)$. Da es nach den eben angestellten Überlegungen für jedes $m \in \mathbb{N}$ eine endliche Folge $g_m \in \ell_{\text{fin}}(\mathbb{C})$ mit $\hat{g}_m J_{\beta_m} = \delta_0 + o_{\ell_2}(\omega^m)$ gibt, gilt gleichfalls

$$J_\varphi = \hat{c}_0 \hat{g}_r J_{\beta_r} + o_{\ell_2}(\omega^r) \, .$$

Das Faltungsprodukt $c := c_0 * g_r \in \ell_{\text{fin}}(\mathbb{C})$ erfüllt also die Behauptung.

Gilt andererseits $J_\varphi = \hat{c} J_{\beta_m} + o_{\ell_2}(\omega^r)$ für eine endliche Folge $c \in \ell_{\text{fin}}(\mathbb{C})$ und ein $m \geq r$, so auch

$$J_\varphi = \hat{c} \left(e_1(-\tfrac{\omega}{2}) \operatorname{sinc}(\omega) \right)^{m+1} \delta_0 + o_{\ell_2}(\omega^r)$$

Analog zu Lemma 4.2.4 findet man eine endliche Folge $c_0 \in \ell_{\text{fin}}(\mathbb{C})$, für welche $\hat{c}_0$ die gleiche Taylorentwicklung bis zum Grad r wie $\hat{c} e_{m+1}(-\frac{\omega}{2}) \operatorname{sinc}(\omega)^{m+1}$ hat. Mit c_0 ist dann die Definition der Approximationsbedingung der Ordnung r erfüllt. □

Korollar 5.3.9 *Seien $m, r \in \mathbb{N}$ mit $m \geq r$, $c \in \ell_{\text{fin}}(\mathbb{C})$ eine endlichen Folge und $\eta \in L^2(\mathbb{R})$ eine Funktion mit beschränkter Prä–Gramscher Faser. Dann hat die Funktion $\varphi = c(\mathcal{T})\beta_m + (1 - \mathcal{T})^{r+1}\eta$ die Approximationsordnung r.*

In umgekehrter Richtung kann von einer Funktion φ mit Approximationsordnung r nur auf die Gestalt $\varphi = c(\mathcal{T})\beta_m + (1 - \mathcal{T})^r \eta$ mit einer endlichen Folge c und einem $\eta \in L^2(\mathbb{R})$ mit Prä–Gramschen Fasern der Ordnung $o_{\ell_2}(\omega^0)$ geschlossen werden. Im Folgenden werden wir aber immer Abschätzungen antreffen, in denen das Landau–Symbol $o_{\ell_2}(\omega^r)$ aus einer Abschätzung $O_{\ell_2}(\omega^{r+1})$ folgt. Eine Funktion φ mit dieser Form der Approximationsbedingung der Ordnung r hat immer die Gestalt

$$\varphi = c(\mathcal{T})\beta_m + (1 - \mathcal{T})^{r+1}\eta \tag{5.6}$$

Lemma 5.3.10 *Seien $s \in \mathbb{N}_{>1}$ ein Skalenfaktor, $a \in \ell_1(\mathbb{C})$ eine Skalierungsfolge und $\varphi \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ eine zulässige Skalierungsfunktion der Approximationsordnung $r \in \mathbb{N}$, welche die Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ erfüllt. Ferner sei das verschiebungsinvariante System zu φ ein Riesz–System, d.h. φ erzeuge eine Multiskalenanalyse.*

Dann gilt für die Fourier–Reihe $\hat{a} = \sum_{n \in \mathbb{Z}} a_n e_{-n}$ zum einen $\hat{a}(0) = s$ und zum anderen hat $\hat{a}$ an den Stellen $\omega = \frac{j}{s}$, $j = 1, \ldots, s-1$, Nullstellen einer Ordnung größer r, d.h. $\hat{a}(\frac{j}{s} + \omega) = 0_{\mathbb{C}}(\omega^r)$.

Beweis: Nach Definition 4.2.5 gibt es ein $c \in \ell_{\text{fin}}(\mathbb{C})$, mit welchem für die Prä–Gramsche Faser $J_\varphi = \hat{c}\delta_0 + o_{\ell_2}(\omega^r)$ gilt.

Die Fourier–Transformierte von φ erfüllt für jedes $\omega \in \mathbb{R}$ die aus der Verfeinerungsgleichung abgeleitete Identität

$$\hat{\varphi}(\omega) = \frac{1}{s}(\mathcal{D}_{\frac{1}{s}}(\hat{a}\hat{\varphi}))(\omega) = \frac{1}{s}\hat{a}(\tfrac{\omega}{s})\hat{\varphi}(\tfrac{\omega}{s}) .$$

Nach Voraussetzung sind sowohl $\hat{a}$ als auch $\hat{\varphi}$ stetig, und es gilt

$$\begin{aligned} s^2\|J_\varphi(\omega) - \hat{c}(\omega)\delta_0\|_{\ell_2}^2 =& s^2 \sum_{n\in\mathbb{Z}} |\hat{\varphi}(n+\omega) - \hat{c}(\omega)\delta_{0,n}|^2 \\ =& \sum_{n\in\mathbb{Z}} \left|\hat{a}(\tfrac{n+\omega}{s})\hat{\varphi}(\tfrac{n+\omega}{s}) - s\hat{c}(\omega)\delta_{0,n}\right|^2 \\ =& \left\|\hat{a}(\tfrac{\omega}{s})\, J_\varphi(\tfrac{\omega}{s}) - s\hat{c}(\omega)\delta_0\right\|_{\ell_2}^2 + \sum_{j=1}^{s-1} |\hat{a}(\tfrac{\omega+j}{s})|^2 \, \|J_\varphi(\tfrac{\omega+j}{s})\|_{\ell_2}^2 . \end{aligned} \tag{5.7}$$

Auf der linken Seite dieser Gleichung steht ein Ausdruck, welcher in Landau–Symbolik die Größe $o_{\mathbb{C}}(\omega^{2r})$ hat. Da die Summanden auf der rechten Seite sämtlich nichtnegativ sind, müssen auch sie diese Größe haben. Dies ist wegen der Riesz–Bedingung an φ nur möglich, wenn $\hat{a}(\frac{j}{s} + w) = o_{\mathbb{C}}(\omega^r)$ für jedes $j = 1, \ldots, s-1$ gilt. Aus der verbleibenden Bedingung von Gleichung (5.7) folgt

$$\hat{a}(\omega)\hat{c}(\omega) - s\hat{c}(s\omega) = o_{\mathbb{C}}(\omega^r) ,$$

aus $\hat{c}(0) = 1$ ergibt sich $\hat{a}(0) = s$. □

Wird $a \in \ell_{\text{fin}}(\mathbb{Z})$ als endliche Folge angenommen, so muss $\hat{a}$ als trigonometrisches Polynom in den Punkten $\frac{1}{s}, \ldots, \frac{s-1}{s}$ jeweils eine Nullstelle vom Grad $r+1$ haben. Für das Laurent–Polynom $a(Z) := \sum_{n\in\mathbb{Z}} a_n Z^n$ bedeutet dies, dass die Linearfaktoren $(Z - e_{\frac{1}{s}}(j))$, $j = 1, \ldots, s-1$, in Vielfachheit $(r+1)$ in der Faktorisierung von $a(Z)$ vorkommen. Es ist

$$(Z - e_{\frac{1}{s}}(1)) \cdots (Z - e_{\frac{1}{s}}(s-1)) = \frac{Z^s - 1}{Z - 1} = 1 + Z + \cdots + Z^{s-1} .$$

Dies ist aber gerade, bis auf einen Faktor $\frac{1}{s}$, das *Haar–Polynom* H_s.

Wegen $\hat{a}(0) = s$ muss es eine Faktorisierung $a(Z) = sH_s(Z)^{r+1}p(Z)$ des Laurent–Polynoms zur Folge a geben, wobei p wieder ein Laurent–Polynom ist und $p(1) = 1$ gilt. Dies entspricht der Forderung, dass der $(s,1)$–periodische Operator $a(\mathcal{T})\,(\uparrow s) : \ell_{\text{fin}}(\mathbb{C}) \to \ell_{\text{fin}}(C)$ eine polynomiale Approximationsordnung A besitzt.

Wir wollen im weiteren auch bei unendlichen Skalierungsfolgen die Struktur $a(Z) = sH_s(Z)^A p(Z)$ mit $A \in \mathbb{N}_{>0}$ und $p \in \ell_1(\mathbb{C})$ voraussetzen. Für $a(Z) = sH_s(Z)^A$ kennen wir schon die B–Splines β_{A-1} als Lösung der zugehörigen Verfeinerungsgleichung.

5.3.4 Biorthogonale und orthogonale Skalierungsfunktionen

In Satz 5.3.4 zur Konstruktion einer Multiskalenanalyse aus einer zulässigen Skalierungsfunktion $\varphi \in L^2(\mathbb{R})$ musste vorausgesetzt werden, dass diese ein verschiebungsinvariantes Riesz–

System erzeugt. Diesen Nachweis für eine Skalierungsfunktion zu führen bzw. diese Eigenschaft aus Eigenschaften der Skalierungsfolge abzuleiten ist im allgemeinen schwierig.

Nimmt man jedoch eine zweite Skalierungsfunktion $\tilde{\varphi} \in L^2(\mathbb{R})$ hinzu, die die Funktion φ im Sinne von Satz 4.2.9 zur Überabtastung ergänzt, so kann dieser Nachweis unter gewissen weiteren Komplementaritätsbedingungen sich als trivial erweisen. Diese Komplementarität ist gesichert, wenn φ und $\tilde{\varphi}$ ein biorthogonales Paar bilden.

Definition 5.3.11 *Seien $\varphi, \tilde{\varphi} \in L^2(\mathbb{R})$ zulässige Skalierungsfunktionen zum Skalenfaktor $s \in \mathbb{N}_{>1}$. Das Paar $(\varphi, \tilde{\varphi})$ wird* biorthogonales Paar *genannt, wenn*

$$\left\langle J_\varphi(\omega),\, J_{\tilde{\varphi}}(\omega) \right\rangle_{\ell_2} = 1$$

fast überall gilt.

Ist das Paar (φ, φ) biorthogonal, so nennt man φ eine orthogonale *Skalierungsfunktion.*

Diese Bedingung ist nicht so einschränkend, wie es scheinen mag; nach Satz 6.3.3 im Anschluss an die Lösungstheorie der Verfeinerungsgleichung werden solche Paare von Skalierungsfunktionen von biorthogonalen Paaren von Skalierungsfolgen erzeugt, sofern diese Skalierungsfolgen auf die im weiteren definierte Art nicht zu groß sind.

Lemma 5.3.12 *Seien $\varphi, \tilde{\varphi} \in L^2(\mathbb{R})$ Funktionen, deren Prä–Gramsche Fasern $J_\varphi, J_{\tilde{\varphi}} : \mathbb{R} \to \ell_2(\mathbb{C})$ essentiell beschränkt sind. Dann gilt*

$$\left\langle J_\varphi(\omega),\, J_{\tilde{\varphi}}(\omega) \right\rangle_{\ell_2} = 1$$

fast überall genau dann, wenn

$$\left\langle \mathcal{T}^n \varphi,\, \tilde{\varphi} \right\rangle = \delta_{0,n} \text{ (Kronecker–Delta)}$$

für alle $n \in \mathbb{Z}$ gilt.

Beweis: Die erste Aussage ist äquivalent dazu, dass für beliebige $c \in \ell_2(\mathbb{C})$ und deren Fourier–Reihe $\hat{c} = \sum_{n\in\mathbb{Z}} c_n e_{-n}$

$$\left\langle \hat{c}\hat{\varphi},\, \hat{\tilde{\varphi}} \right\rangle_{L^2} = \int_0^1 \left\langle \hat{c}(\omega) J_\varphi(\omega),\, J_{\tilde{\varphi}}(\omega) \right\rangle_{\ell_2} \mathrm{d}\omega = \int_0^1 \hat{c}(\omega)\, \mathrm{d}\omega = c_0$$

gilt. Nach den Rechenregeln der Fourier–Transformation und dem Satz von Plancherel gilt weiter

$$c_0 = \left\langle \hat{c}\hat{\varphi},\, \hat{\tilde{\varphi}} \right\rangle_{L^2} = \left\langle c(\mathcal{T})\varphi,\, \tilde{\varphi} \right\rangle_{L^2} .$$

Dass dies für beliebige $c \in \ell_2(\mathbb{C})$ gilt, ist äquivalent zur zweiten Aussage. □

Eine andere Formulierung dieser Eigenschaft ist, dass die Kombination des Synthese–Operators $\mathcal{E}_\varphi$ mit dem Analyse–Operator $\mathcal{E}^*_{\tilde{\varphi}}$ die Identität auf $\ell_2(\mathbb{C})$ ergibt, denn

$$\mathcal{E}^*_{\tilde{\varphi}}\mathcal{E}_\varphi(c) = \mathcal{E}^*_{\tilde{\varphi}}\Big(\sum_{n\in\mathbb{Z}} c_n \mathcal{T}^n \varphi\Big) = \Big\{\sum_{n\in\mathbb{Z}} c_n \left\langle \mathcal{T}^n\varphi,\, \mathcal{T}^k\tilde{\varphi} \right\rangle\Big\}_{k\in\mathbb{Z}} = c\,.$$

Satz 5.3.13 *Seien $\varphi, \tilde{\varphi} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ zulässige Skalierungsfunktionen zu einem gemeinsamen Skalenfaktor $s \in \mathbb{N}_{>1}$, die ein biorthogonales Paar $(\varphi, \tilde{\varphi})$ bilden. Dann erzeugen sowohl φ als auch $\tilde{\varphi}$ eine Multiskalenanalyse.*

Beweis: Da sowohl φ als auch $\tilde{\varphi}$ eine Approximationsbedingung erfüllen, erzeugen beide je ein verschiebungsinvariantes Bessel–System. Um die Voraussetzungen des Satzes 5.3.4 zu erfüllen, müssen diese auch Riesz–Systeme sein. Sei $c \in \ell_2(\mathbb{C})$ beliebig. Dann gilt für das Skalarprodukt der mit c gebildeten Funktionenreihen

$$\left\langle \mathcal{E}_\varphi(c), \mathcal{E}_{\tilde{\varphi}}(c) \right\rangle_{L^2} = \left\langle \mathcal{E}_{\tilde{\varphi}}^* \mathcal{E}_\varphi(c), c \right\rangle_{L^2} = \|c\|_{\ell_2}^2 .$$

Andererseits gilt für das Skalarprodukt die Cauchy–Schwarzsche Ungleichung, und mit der Schranke $\sqrt{\tilde{B}}$ für die Abbildung $\mathcal{E}_\varphi$ folgt

$$\left\langle \mathcal{E}_\varphi(c), \mathcal{E}_{\tilde{\varphi}}(c) \right\rangle_{L^2} \leq \|\mathcal{E}_\varphi(c)\|_{L^2} \|\mathcal{E}_{\tilde{\varphi}}(c)\|_{L^2} \leq \|\mathcal{E}_\varphi(c)\|_{L^2} \sqrt{\tilde{B}} \|c\|_{\ell_2} .$$

Zusammengefasst gilt also für beliebiges $c \in \ell_2(\mathbb{C})$

$$\frac{1}{\tilde{B}} \|c\|_{\ell_2}^2 \leq \|\mathcal{E}_\varphi(c)\|_{L^2}^2 .$$

Somit ist $\{\mathcal{T}^n \varphi : n \in \mathbb{Z}\}$ ein Riesz–System und φ erzeugt eine Multiskalenanalyse. Analog argumentiert man für $\tilde{\varphi}$. □

Die Funktionen $\varphi, \tilde{\varphi} \in L^2(\mathbb{R})$ bilden nach Lemma 5.3.12 genau dann ein biorthogonales Paar, wenn $\mathcal{E}_{\tilde{\varphi}}^* \mathcal{E}_\varphi = id_{\ell_2(\mathbb{C})}$ gilt. Für den Synthese–Operator zu φ erhalten wir unter Benutzung der Verfeinerungsgleichung

$$\begin{aligned} \mathcal{E}_\varphi(c) &= \sum_{n \in \mathbb{Z}} c_n \mathcal{T}^n \varphi = \sum_{n \in \mathbb{Z}} c_n \mathcal{T}^n \mathcal{D}_s a(\mathcal{T}) \varphi \\ &= \mathcal{D}_s a(\mathcal{T}) \sum_{n \in \mathbb{Z}} c_n \mathcal{T}^{sn} \varphi = \mathcal{D}_s \mathcal{E}_\varphi \left(a(\mathcal{T}) \left(\uparrow \mathrm{s}\right)(c) \right) . \end{aligned}$$

Analog dazu ergibt sich $\mathcal{E}_{\tilde{\varphi}} = \mathcal{D}_s \mathcal{E}_{\tilde{\varphi}} \circ a(\mathcal{T}) \circ \left(\uparrow \mathrm{s}\right)$. Dies zusammensetzend ergibt sich

$$id_{\ell_2(\mathbb{C})} = \left(\downarrow \mathrm{s}\right) \tilde{a}(\mathcal{T})^* \mathcal{E}_{\tilde{\varphi}}^* \mathcal{D}_s^* \mathcal{D}_s \mathcal{E}_\varphi a(\mathcal{T}) \left(\uparrow \mathrm{s}\right) = \tfrac{1}{s} \left(\downarrow \mathrm{s}\right) \tilde{a}(\mathcal{T})^* a(\mathcal{T}) \left(\uparrow \mathrm{s}\right) .$$

Weiter ist $a(\mathcal{T}) \left(\uparrow \mathrm{s}\right)$ ein $(s,1)$–periodischer linearer Operator und hat die Polyphasendarstellung

$$a(\mathcal{T}) \left(\uparrow \mathrm{s}\right) = \left(\uparrow\uparrow \mathrm{s}\right) \begin{pmatrix} a_{(0)}(\mathcal{T}) \\ a_{(1)}(\mathcal{T}) \\ \vdots \\ a_{(s-1)}(\mathcal{T}) \end{pmatrix}$$

mit den Polyphasenteilfolgen $a_{(k)} = \{a_{k+sn}\}_{n \in \mathbb{Z}}$ und deren Differenzenoperatoren $a_{(k)}(\mathcal{T}) = \sum_{n \in \mathbb{Z}} a_{k+sn} \mathcal{T}^n$. Mit der analogen Zerlegung von $\tilde{a}$ kann die oben erhaltene Identität für die Skalierungsfolgen eines biorthogonalen Paares auch in die Form

$$s\, id_{\ell_2(\mathbb{C})} = \left(\downarrow \mathrm{s}\right) \tilde{a}(\mathcal{T})^* a(\mathcal{T}) \left(\uparrow \mathrm{s}\right) = \sum_{k=0}^{s-1} \tilde{a}_{(k)}(\mathcal{T})^* a_{(k)}(\mathcal{T})$$

gebracht werden. Dass aus dieser Identität unter bestimmten Umständen wiederum auf die Biorthogonalität von $(\varphi, \tilde{\varphi})$ geschlossen werden kann, wird im Anschluss an die Lösungstheorie der Skalierungsgleichung in Abschnitt 6.3 gezeigt.

5.3.5 Wavelet–Systeme

Sei $\varphi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ eine zulässige Skalierungsfunktion, welche eine Multiskalenanalyse $\{V_j\}_{j\in\mathbb{Z}}$ des $L^2(\mathbb{R})$ erzeugt. Zu dieser soll, analog zu den Haar– und Kardinalwavelets, eine weiter Folge $\{W_j\}_{j\in\mathbb{Z}}$ von Unterräumen gefunden werden, so dass es für beliebige ganze Zahlen $K > J$ eine Zerlegung

$$V_K = V_J \oplus W_J \oplus W_{J+1} \oplus \cdots \oplus W_{K-1}$$

gibt. Die direkten Summen brauchen nicht orthogonal zu sein; die Unterräume sollen jedoch wieder gestreckte oder gestauchte Kopien voneinander sein, d.h. für jedes $j \in \mathbb{Z}$ soll $W_j = \mathcal{D}_s^j W_0$ gelten.

W_0 soll ein endlich erzeugter verschiebungsinvarianter Raum sein, d.h. von endlich vielen Funktionen $\Psi' = \{\psi_2, \ldots, \psi_M\} \subset V_1$ und deren ganzzahlig verschobenen Kopien aufgespannt werden, $W_0 = \operatorname{span} X(\Psi')$. Wir schwächen die Bedingung $V_1 = V_0 \oplus W_0$ dahingehend ab, dass das von $\psi_1 := \varphi, \psi_2, \ldots, \psi_M$ erzeugte verschiebungsinvariante System $X(\Psi)$, $\Psi = \{\psi_k : k = 1, \ldots, M\}$, ein Frame in V_1 sein soll. Die Funktionen in $\psi_2, \ldots, \psi_M$ werden *Mutter–Wavelets* oder einfach *Wavelets* genannt.

Ist ein solch komplementärer Unterraum W_0 gefunden, so können wir für jedes $j \in \mathbb{Z}$ diesen um den Faktor s^j strecken und erhalten einen Unterraum $W_j = \mathcal{D}_s^j W_0$, welcher komplementär zu V_j in V_{j+1} ist. Die Struktur der Gesamtheit der Erzeugendensysteme dieser komplementären Räume wird *affines System* genannt.

Definition 5.3.14 *Seien $\Psi' \subset L^2(\mathbb{R})$ eine endliche Teilmenge und $s \in \mathbb{N}_{>1}$ ein Skalenfaktor. Das von Ψ' erzeugte* affine System *ist die Menge*

$$\operatorname{Aff}_s(\Psi') := \left\{ s^{\frac{j}{2}} \mathcal{D}_s^j \mathcal{T}^n \psi : \psi \in \Psi',\ j, n \in \mathbb{Z} \right\} .$$

Ein affines System, welches ein Frame in $L^2(\mathbb{R})$ ist, wird Wavelet–System *genannt.*

Sei also die Multiskalenanalyse von einer zulässigen Skalierungsfunktion $\varphi \in L^1 \cap L^2$ erzeugt. Dann sind die Elemente von V_1 Linearkombinationen $\sum_{n\in\mathbb{Z}} c_n \mathcal{T}^n \varphi$ mit $c = \{c_n\} \in \ell_2(\mathbb{C})$. Wir schränken uns auf den Fall ein, in welchem die Wavelet–Funktionen mittels Folgen aus $\ell_1(\mathbb{C})$ darstellbar sind. Seien, neben der Skalierungsfolge $b_1 := a$ von $\psi_1 := \varphi$, weitere Folgen $b_2, \ldots, b_M \in \ell_1(\mathbb{Z})$ gewählt, und mit diesen die Differenzenoperatoren

$$b_k(\mathcal{T}) := \sum_{n\in\mathbb{Z}} b_{k,n} \mathcal{T}^n \text{ und die Funktionen } \psi_k := \mathcal{D}_s b_k(\mathcal{T}) \varphi,$$

$k = 1, \ldots, M$, konstruiert. Der Synthese–Operator $\mathcal{E}_\Psi : \ell_2(\mathbb{C}^M) \to V_1 \in L^2(\mathbb{R})$ zu $\Psi = \{\psi_1, \ldots, \psi_M\}$ ist nach diesen Voraussetzungen beschränkt. Für jedes Tupel $d = (d_1, \ldots, d_M)^t \in$

$\ell_2(\mathbb{C}^M)$ von Koeffizientenfolgen gilt

$$\begin{aligned}\mathcal{E}_\Psi(d) = \sum_{k=1}^{M}\sum_{n\in\mathbb{Z}} d_{k,n}\mathcal{T}^n\psi_k &= \mathcal{D}_s\left(\sum_{k=1}^{M}\sum_{n\in\mathbb{Z}} d_{k,n}\mathcal{T}^{sn}b_k(\mathcal{T})\right)\varphi\\ &= \mathcal{D}_s\mathcal{E}_\varphi\left(\sum_{k=1}^{M} b_k(\mathcal{T})\left(\uparrow\mathrm{s}\right)d_k\right)\end{aligned}$$

Soll $X(\Psi)$ ein Frame von V_1 sein, muss notwendigerweise $\mathcal{E}_\Psi$ surjektiv auf V_1 sein. Dies ist gewährleistet, wenn der $(s,1)$–periodische lineare Operator

$$B := (b_1(\mathcal{T}),\ldots,b_m(\mathcal{T}))\circ\left(\uparrow\mathrm{s}\right) : \ell_2(\mathbb{C}^M) \to \ell_2(\mathbb{C})$$

surjektiv ist. Ist beispielsweise der adjungierte Operator $\sqrt{s^{-1}}B^* : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C}^M)$ semi–unitär, so ist für eine beliebige Folge $c \in \ell_2(\mathbb{C})$ die vektorwertige Folge $d := \frac{1}{s}B^*c$ eine Lösung des linearen Gleichungssystems $c = B(d)$.

Ohne Beweis zitieren wir

Satz 5.3.15 (s. [RS97b, RS97a])
Sei $\varphi \in L^2(\mathbb{R})$ eine zulässige Skalierungsfunktion zum Skalenfaktor $s \in \mathbb{N}_{>1}$ und mit Skalierungsfolge $a \in \ell_1(\mathbb{Z})$. Es sei weiter vorausgesetzt, dass das von φ erzeugte verschiebungsinvariante System ein Riesz–System ist, d.h. φ eine Multiskalenanalyse erzeugt.

Seien $b_1 := a$ und $b_2,\ldots,b_M \in \ell_1(\mathbb{Z})$ weitere Folgen. Mit diesen seien die Differenzenoperatoren $b_k(\mathcal{T}) := \sum_{n\in\mathbb{Z}} b_{k,n}\mathcal{T}^n$ gebildet. Für den $(s,1)$–periodischen linearen Operator $B : \ell_2(\mathbb{C}^M) \to \ell(\mathbb{C})$,

$$\ell_2(\mathbb{C}^M) \ni d \mapsto B(d) := b_1(\mathcal{T})\left(\uparrow\mathrm{s}\right)(d_1) + \cdots + b_M(\mathcal{T})\left(\uparrow\mathrm{s}\right)d_M$$

gelte die Identität $BB^* = s\,id_{\ell_2(\mathbb{C})}$, d.h. $\sqrt{s^{-1}}B^*$ sei semi–unitär.

Dann erzeugen die Funktionen $\psi_2,\ldots,\psi_M \in L^2(\mathbb{R})$,

$$\psi_i := \mathcal{D}_s b_i(\mathcal{T})\varphi = \sum_{n\in\mathbb{Z}} b_{i,n}\mathcal{D}_s\mathcal{T}^n\psi$$

ein affines System, welches ein straffer Frame von $L^2(\mathbb{R})$ und damit ein Wavelet–System ist.

Als $(s,1)$–periodischer Operator besitzt B eine Darstellung $B = \left(\uparrow\uparrow\mathrm{s}\right)\circ B_{poly}$ mit der Polyphasenmatrix

$$B_{poly} = \begin{pmatrix} a_{(0)}(\mathcal{T}) & b_{2,(0)}(\mathcal{T}) & \ldots & b_{M,(0)}(\mathcal{T})\\ \vdots & \vdots & & \vdots\\ a_{(s-1)}(\mathcal{T}) & b_{2,(s-1)}(\mathcal{T}) & \ldots & b_{M,(s-1)}(\mathcal{T})\end{pmatrix},$$

wobei die Polyphasen durch $(b_{k,(0)},\ldots,b_{k,(s-1)}) := \left(\downarrow\downarrow\mathrm{s}\right)b_k$ gegeben sind, d.h. deren Differenzenoperatoren sind

$$b_{k,(m)}(\mathcal{T}) := \sum_{n\in\mathbb{Z}} b_{k,m+sn}\mathcal{T}^n, \quad k = 1,\ldots,M, \quad m = 0,\ldots,s-1\,.$$

Ist die Skalierungsfunktion φ orthogonal und ihre Skalierungsfolge a endlich, so ist

$$\sqrt{s^{-1}}(a_{(0)}(\mathcal{T}),\ldots,a_{(s-1)}(\mathcal{T}))^t : \ell_{\text{fin}}(\mathbb{C}) \to \ell_{\text{fin}}(\mathbb{C}^s)$$

eine semi–unitäre Abbildung. Diese kann nach den Bemerkungen zu Satz 3.6.2 zu einer unitären Abbildung $\sqrt{s^{-1}}B : \ell_{\text{fin}}(\mathbb{C}^s) \to \ell_{\text{fin}}(\mathbb{C}^s)$ ergänzt werden, woraus sich die $s-1$ endlichen Folgen $b_2,\ldots,b_s \in \ell_{\text{fin}}(\mathbb{C})$ rekonstruieren lassen. Die daraus konstruierten Wavelets $\psi_2,\ldots,\psi_s$ erzeugen ein verschiebungsinvariantes Orthonormalsystem $X(\varphi,\psi_2,\ldots,\psi_s)$. Daraus folgt, dass der Raum $W_0 = \text{span}(X(\psi_2,\ldots,\psi_s))$ senkrecht zu $V_0 = \text{span}(X(\varphi))$ steht, daher alle Komplemente $W_j = \mathcal{D}_s^j W_0$ paarweise senkrecht zueinander stehen und $\text{Aff}(\psi_2,\ldots,\psi_s)$ gleichzeitig Orthonormalsystem und Frame, also eine *Hilbert–Basis* ist.

Ist die Skalierungsfunktion φ nicht orthogonal, so ist trotzdem der zu ihrer Skalierungsfolge a bzw. deren Differenzenoperator $a(\mathcal{T})$ gebildete Operator

$$T := \big(\downarrow\text{s}\big)\, a(\mathcal{T})^* a(\mathcal{T})\, \big(\uparrow\text{s}\big) = \sum_{k=0}^{s-1} \tilde{a}_{(k)}(\mathcal{T})^* a_{(k)}(\mathcal{T})$$

selbstadjungiert und positiv semidefinit. Ist dieser Operator durch $\sqrt{s}$ beschränkt, d.h. gilt

$$\|a(\mathcal{T})\,\big(\uparrow\text{s}\big)(c)\|_{\ell_2}^2 = \sum_{k=0}^{s-1} \|a_{(k)}(\mathcal{T})c\|_{\ell_2}^2 \le s\|c\|_{\ell_2}^2$$

für alle $c \in \ell_2(\mathbb{C})$, so ist die Differenz $(s\ id_{\ell_2} - T)$ ebenfalls selbstadjungiert und positiv semidefinit. Es kann dann mittels des Fejer–Riesz–Algorithmus (s. [PS71], nach [BKN94]) eine endliche Folge r gefunden werden, so dass mit deren Differenzenoperator $r(\mathcal{T})$

$$T + r(\mathcal{T})^* r(\mathcal{T}) = \sum_{k=0}^{s-1} \tilde{a}_{(k)}(\mathcal{T})^* a_{(k)}(\mathcal{T}) + r(\mathcal{T})^* r(\mathcal{T}) = s\, id_{\ell_2}$$

gilt. D.h. die Abbildung

$$\sqrt{s^{-1}}(a_{(0)}(\mathcal{T}),\ldots,a_{(s-1)}(\mathcal{T}),r(\mathcal{T}))^t : \ell_{\text{fin}}(\mathbb{C}) \to \ell_{\text{fin}}(\mathbb{C}^{s+1})$$

ist semi–unitär. Somit kann diese Spalte wieder zu einer quadratischen unitären Operatormatrix $\sqrt{s^{-1}}\tilde{B} : \ell_{\text{fin}}(\mathbb{C}^{s+1}) \to \ell_{\text{fin}}(\mathbb{C}^{s+1})$ ergänzt werden. Deren Einträge interpretieren wir als

$$\tilde{B} = \begin{pmatrix} a_{(0)}(\mathcal{T}) & b_{2,(0)} & \ldots & b_{s+1,(0)} \\ \vdots & \vdots & & \vdots \\ a_{(s-1)}(\mathcal{T}) & b_{2,(s-1)} & \ldots & b_{s+1,(s-1)} \\ r_1(\mathcal{T}) & r_2(\mathcal{T}) & \ldots & r_{s+1}(\mathcal{T}) \end{pmatrix}$$

Streichen wir von dieser Matrix die letzte Zeile, so erhalten wir einen Operator $B_{poly} : \ell_{\text{fin}}(\mathbb{C}^{s+1}) \to \ell_{\text{fin}}(\mathbb{C}^s)$, dessen erste Spalte die Polyphasen von $a(\mathcal{T})$ enthält und dessen adjungierter Operator $\sqrt{s^{-1}}B^*$ semi–unitär ist. Die mit den weiteren Spalten von B gebildeten s Wavelets $\psi_2,\ldots,\psi_{s+1}$ erzeugen somit einen *Wavelet-Frame*.

Praktisch wichtig sind auch die *biorthogonalen Wavelet–Systeme*. Als biorthogonales Wavelet–System bezeichnet man ein Paar von affinen Systemen, von welchen jedes ein Frame in $L^2(\mathbb{R})$ ist und jeweils das eine der duale Frame des anderen affinen Systems sind.

Satz 5.3.16 (s. [RS97b, RS97a] (ohne Beweis))
Seien $\varphi, \tilde{\varphi}$ zulässige Skalierungsfunktionen zum Skalenfaktor $s \in \mathbb{N}_{>1}$ mit Skalierungsfolgen $b_1 := a, \tilde{b}_1 := \tilde{a} \in \ell_1(\mathbb{C})$. Das von diesen erzeugte verschiebungsinvariante System sei jeweils ein Riesz–System.

Es seien weitere Folgen $b_2, \ldots, b_M, \tilde{b}_2, \ldots, \tilde{b}_M \in \ell_1(\mathbb{C})$ gewählt. Sind die Operatoren

$$B := (b_1(\mathcal{T}), \ldots, b_M(\mathcal{T}))\,(\uparrow \mathrm{s}) : \ell_2(\mathbb{C}^M) \to \ell_2(\mathbb{C})$$
und
$$\tilde{B} := (\tilde{b}_1(\mathcal{T}), \ldots, \tilde{b}_M(\mathcal{T}))\,(\uparrow \mathrm{s}) : \ell_2(\mathbb{C}^M) \to \ell_2(\mathbb{C})$$

zueinander komplementär in dem Sinne, dass $B\tilde{B}^* = s\,id_{\ell_2}$ gilt, dann erzeugen die Funktionen $\psi_2, \ldots, \psi_M \in L^2(\mathbb{R})$,

$$\psi_i := \mathcal{D}_s b_i(\mathcal{T})\varphi = \sum_{n\in\mathbb{Z}} b_{i,n}\mathcal{D}_s\mathcal{T}^n\psi$$

bzw. $\tilde{\psi}_2, \ldots, \tilde{\psi}_M \in L^2(\mathbb{R})$

$$\tilde{\psi}_i := \mathcal{D}_s b_i(\mathcal{T})\tilde{\varphi} = \sum_{n\in\mathbb{Z}} b_{i,n}\mathcal{D}_s\mathcal{T}^n\tilde{\psi}$$

affine Systeme, welche zueinander duale Wavelet–Systeme sind.

Diese Konstruktion ist selbst bei endlichen Skalierungsfolgen unter allgemeinsten Voraussetzungen ausführbar. Ist $(\varphi, \tilde{\varphi})$ sogar ein biorthogonales Paar, so kann die Anzahl der *Mutter–Wavelets* auf $s-1$ beschränkt werden, d.h. es kann $M = s$ realisiert werden.

Kapitel 6

Verfeinerungsgleichung und Skalierungsfunktion

Wir wollen Bedingungen angeben, unter welchen die Lösung der Verfeinerungsgleichung (5.3) zu einer gegebenen Skalierungsfolge eindeutig ist und die in Satz 5.3.4 angegebenen Bedingungen erfüllt. Die dazu benutzten Techniken lehnen sich an die klassische Methode an, die Verfeinerungsgleichung als iteriertes Funktionensystem zu betrachten (s. [Dau92]).

Betrachten wir als Beispiel den Faktor $s = 2$ und und eine Skalierungsfolge

$$a = (\dots, 0, a_0, \dots, a_{2N-1}, 0, \dots), \qquad N \in \mathbb{N}\,,$$

welche endlich ist. Gibt es eine Lösung der Verfeinerungsgleichung $\varphi = \mathcal{D}_2 a(\mathcal{T})\varphi$ mit kompaktem Träger, so muss dieser im Intervall $[0, 2N-1]$ enthalten sein.

Jeder Funktion f mit Träger in $[0, 2N-1]$ kann eine vektorwertige Funktion $\mathbf{v} : [0,1] \to \mathbb{R}^{2N-1}$ zugeordnet werden, welche komponentenweise als $\mathbf{v}_k(x) := f(k+x)$, $k = 0, 1, \dots, 2N-2$, gegeben ist. Wir können diese Zuordnung in die Verfeinerungsgleichung einsetzen, für jedes $x \in \left[0, \frac{1}{2}\right)$ ist

$$\mathbf{v}_k(x) = f(k+x) = \sum_{n\in\mathbb{Z}} a_n f(2x + 2k - n) = \sum_{m=0}^{2N-2} a_{2k-m}\mathbf{v}_m(2x)\,.$$

Analog ergibt sich eine solche Identität für $x \in [\frac{1}{2}, 1)$, wir erhalten für die beiden Hälften des Intervalls jeweils eine Transformationsmatrix,

$$\mathbf{v}(x) = \begin{cases} A_0\mathbf{v}(2x) & x \in [0, \frac{1}{2}) \\ A_1\mathbf{v}(2x-1) & x \in [\frac{1}{2}, 1] \end{cases}\,.$$

Dabei sind die Matrizen A_0, A_1 Block–Toeplitz–Matrizen der Dimension $2N-1$,

$$A_0 = \begin{pmatrix} a_0 & 0 & \dots & 0 & 0 \\ a_2 & a_1 & \dots & 0 & 0 \\ \vdots & & \vdots & & \vdots \\ a_{2N-2} & a_{2N-3} & \dots & a_1 & a_0 \\ 0 & a_{2N-1} & \dots & a_3 & a_2 \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & a_{2N-1} & a_{2N-2} \end{pmatrix}, \quad A_1 = \begin{pmatrix} a_1 & a_0 & \dots & 0 & 0 \\ a_3 & a_2 & \dots & 0 & 0 \\ \vdots & & \vdots & & \vdots \\ a_{2N-1} & a_{2N-2} & \dots & a_2 & a_1 \\ 0 & 0 & \dots & a_4 & a_3 \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & a_{2N-1} \end{pmatrix}\,.$$

Es ergibt sich aus der allgemeinen Lösungstheorie solcher Selbstähnlichkeitsbeziehungen (s. [CM99]), dass nur dann auf die Existenz einer Lösung geschlossen werden kann, wenn

beide Matrizen A_0, A_1 bzgl. einer Norm auf dem $\mathbb{R}^{2N-1}$ kontraktiv sind. In diesem Fall jedoch ist die Lösung die Nullfunktion.

Um nichttriviale Lösungen zu erhalten, werden die Werte von $\mathbf{v}$ auf einen affinen Unterraum eingeschränkt, und dementsprechend die Kontraktivität der linearen Abbildungen auf diesen Unterräumen untersucht. Sei dazu vorausgesetzt, dass $\varphi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ eine Approximationsbedingung erfüllt. Dann müssen notwendigerweise $\hat{a}(0) = a(1) = 2$ sowie $\hat{a}(\frac{1}{2}) = a(-1) = 0$ gelten. Das Polynom $a(Z) = a_0 + a_1 Z + \cdots + a_{2N-1} a^{2N-1}$ hat einen Faktor $(1+Z)$. Dann ergibt die Summe $\mathbf{v}_0(x) + \cdots + \mathbf{v}_{2N-2}(x) = \sum_{n\in\mathbb{Z}} \varphi(x+n)$ fast überall dieselbe Konstante. Da die Verfeinerungsgleichung homogen ist, kann diese Konstante als 1 gewählt werden.

Für den Raum der endlichen Folgen mit Summe 1 kann der Ansatz $\mathbf{v}(x) = \delta^0 + (1-\mathcal{T})\mathbf{w}(x)$, d.h. $\varphi = \chi_{[0,1)} + (1-\mathcal{T})\eta$ gemacht werden, wobei $\mathbf{w} : [0,1] \to \ell_{\text{fin}}(\mathbb{R})$ eine weitere Abbildung in den Raum der endlichen Folgen und $\eta \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ die aus diesen Stücken zusammengesetzte Funktion ist. Die Skalierungsfolge kann als $a(Z) = (1+Z)p(Z)$ faktorisiert werden, die Verfeinerungsgleichung transformiert sich damit zu

$$\begin{aligned}
\chi_{[0,1)} + (1-\mathcal{T})\eta = \varphi &= \mathcal{D}_2 a(\mathcal{T})\varphi \\
&= \mathcal{D}_2(1+\mathcal{T})p(\mathcal{T})\chi_{[0,1)} + (1-\mathcal{T})\mathcal{D}_2 p(\mathcal{T})\eta \\
\iff (1-\mathcal{T})\eta &= \mathcal{D}_2(1+\mathcal{T})(p(\mathcal{T})-1)\chi_{[0,1)} + (1-\mathcal{T})\mathcal{D}_2 p(\mathcal{T})\eta\,.
\end{aligned}$$

Da nun aber $p(1) = 1$ gelten muss, gibt es eine endliche Folge $\tilde{c}$ mit $p(Z) - 1 = (1-Z)\tilde{c}(Z)$. Aus der oben angegebenen Gleichung kann der gemeinsame Faktor $(1-\mathcal{T})$ entfernt werden, da alle auftretenden Funktionen einen endlichen Träger besitzen. Die resultierende Gleichung

$$\eta = \mathcal{D}_2\tilde{c}(\mathcal{T})\chi_{[0,1)} + \mathcal{D}_2 p(\mathcal{T})\eta$$

ist affin linear und kann mittels des Vektors $\mathbf{w}$ der Funktionswerte von η in der Form

$$\mathbf{w}(x) = \begin{cases} \tilde{C}_0 + P_0\mathbf{w}(2x) & x \in [0, \frac{1}{2}) \\ \tilde{C}_1 + P_1\mathbf{w}(2x-1) & x \in [\frac{1}{2}, 1] \end{cases}$$

mit Vektoren $\tilde{C}_0$ und $\tilde{C}_1$ sowie Matrizen P_0, P_1 nach dem Muster von A_0 und A_1 angegeben werden. Sind diese Matrizen bzw. der Operator $\mathcal{D}_2 p(\mathcal{T})$ kontraktiv bzgl. einer Norm, so hat dieses System affin linearer Gleichungen einen von Null verschiedenen Fixpunkt. Die Qualität des Fixpunkts hängt von der Norm ab, für welche die Kontraktivität nachgewiesen wurde.

Diese Methode wurde schon in [Dau92] dargestellt und in [CM99] in einen allgemeineren Kontext der Selbstähnlichkeit gestellt. Jedoch wurden die notwendigen Einschränkungen auf affine Unterräume dort nur im Nachhinein und am Einzelfall betrachtet. Im folgenden werden wir diese Reduktion auf einen Unterraum als zentralen Teil der Existenztheorie einführen. Die daran anschließende Lösungstheorie inhomogener Verfeinerungsgleichungen wird schon in [SZ98], jedoch nicht im Zusammenhang mit dem hier zu betrachtenden Problem diskutiert.

6.1 Zur Lösbarkeit der Verfeinerungsgleichung

Wir werden zunächst ein allgemeines Resultat zur Existenz von Skalierungsfunktionen bei gegebener Skalierungsfolge beweisen. Dabei ist die Norm auf dem Funktionenraum, bis auf einige recht allgemeine Forderungen, beliebig. Im nachfolgenden Abschnitt kann dann aus der Lösbarkeit der Verfeinerungsgleichung für spezielle Normen auf analytische Eigenschaften der Lösung geschlossen werden.

6.1.1 Schnell fallende Folgen

Um die Lösbarkeit der Verfeinerungsgleichung nachzuweisen, benötigen wir, dass die Skalierungsfolge sich „fast“ wie eine endliche Folge verhält.

Definition 6.1.1 *Für jedes $M \in \mathbb{N}_{>0}$ bezeichnen wir mit*

$$\ell_{1,M}(\mathbb{C}) := \ell_{1,M}(\mathbb{Z}, \mathbb{C})$$

diejenige Teilmenge des Folgenraums $\ell_1(\mathbb{Z})$, welche alle Elemente $c \in \ell_1(\mathbb{Z})$ enthält, die $\sum_{n\in\mathbb{Z}} |c_n|\,|n|^M < \infty$ erfüllen.

Wie man sich leicht überzeugt, ist $\ell_{1,M}(\mathbb{Z})$ ein Untervektorraum von $\ell_1(\mathbb{Z})$. Mit der Norm $\|c\|_{1,M} := \sum_{n\in\mathbb{Z}}(1+|n|)^M|c_n|$ ist $\ell_{1,M}(\mathbb{Z})$ ein Banachraum, wegen $1+|n+1| \le 2+2|n|$ ist 2^M eine Schranke für den Translationsoperator. Insbesondere definiert jedes Laurent–Polynom einen beschränkten Operator auf $\ell_{1,M}(\mathbb{Z})$.

Lemma 6.1.2 *Zu jeder Folge $p \in \ell_{1,M}(\mathbb{C})$ gibt es eine endliche Folge $q \in \ell_{\text{fin}}(\mathbb{C})$ und einen „Rest“ $r \in \ell_1(\mathbb{C})$, so dass*

$$p = q + (1-\mathcal{T})^M r$$

gilt. Sind $q_1, q_2 \in \ell_{\text{fin}}(\mathbb{C})$ und $r_1, r_2 \in \ell_1(\mathbb{C})$ zwei verschiedene Lösungen dieses Zerlegungsproblems, so gibt es eine endliche Folge $d \in \ell_{\text{fin}}(\mathbb{C})$ mit $q_2 = q_1 + (1-\mathcal{T})^M d$ und dementsprechend $r_2 = r_1 - d$.

Beweis: Wir zeigen diese Behauptung schrittweise. Sei $p \in \ell_{1,m}(\mathbb{C})$ mit $m \ge 1$, dann gibt es ein $r \in \ell_{1,m-1}(\mathbb{C})$, so dass mit $\tilde{p} := \sum_{n\in\mathbb{Z}} p_n \in \mathbb{C}$ die Identität

$$p = \tilde{p}\delta^0 + (1-\mathcal{T})r$$

gilt. Diese Gleichung kann direkt gelöst werden, es gilt

$$\begin{aligned} p - \tilde{p}\delta^0 &= \sum_{n\in\mathbb{Z}} p_n(\delta^n - \delta^0) = \sum_{n\in\mathbb{Z}} p_n(\mathcal{T}^n - 1)\delta^0 \\ &= \sum_{n=1}^{\infty} p_n(\mathcal{T}-1)\sum_{k=0}^{n-1}\mathcal{T}^k\delta^0 + \sum_{n=-\infty}^{-1} p_n(1-\mathcal{T})\sum_{k=n}^{-1}\mathcal{T}^k\delta^0 \\ &= (1-\mathcal{T})\left(-\sum_{0\le k<n} p_n\delta^k + \sum_{n\le k<0} p_n\delta^k\right). \end{aligned}$$

Daraus liest man die Koeffizienten der „Restfolge“ r ab zu

$$r_k = \begin{cases} -\sum_{n=k+1}^{\infty} p_n & \text{bei } k \ge 0 \\ \sum_{n=-\infty}^{k} p_n & \text{bei } k < 0\,. \end{cases}$$

Wegen $p \in \ell_{1,m}(\mathbb{C})$ sind diese Reihen absolut konvergent. Es verbleibt $r \in \ell_{1,m-1}(\mathbb{C})$ nachzuweisen. Für die entsprechende Norm gilt

$$\|r\|_{1,m-1} = \sum_{k\in\mathbb{Z}} |k|^{m-1}\,|r_k| \le \sum_{0\le k<n} |k|^{m-1}\,|p_n| + \sum_{n\le k<0} |k|^{m-1}\,|p_n| \le \sum_{n\in\mathbb{Z}} |n|^m\,|p_n| < \infty$$

Wenden wir diese Konstruktion mehrfach auf r anstelle von p an, so erhalten wir die behauptete Zerlegung.

Für zwei verschiedene Zerlegungen $p = q_1 + (1-\mathcal{T})^M r_1 = q_2 + (1-\mathcal{T})^M r_2$ gilt

$$q_2 - q_1 = (1-\mathcal{T})^M (r_1 - r_2)\,.$$

Da die linke Seite endlich ist, und die Differenz $r_1 - r_2 \in \ell_1(\mathbb{C})$ absolut summierbar, muss schon diese Differenz eine endliche Folge $d := r_1 - r_2 \in \ell_{\text{fin}}(\mathbb{C})$ sein, d.h. es gilt $q_2 = q_1 + (1-\mathcal{T})^M d$. Denn die Gleichung $a = (1-\mathcal{T})b$ mit einer endlichen Folge a hat Lösungen b, welche in beiden Richtungen gegen Unendlich stationär werden. Somit kann $b \in \ell_1(\mathbb{C})$ nur gelten, wenn $b \in \ell_{\text{fin}}(\mathbb{C})$ eine endliche Folge ist. Per Induktion gilt dasselbe für $a = (1-\mathcal{T})^M b$ mit $M > 0$ beliebig und $a \in \ell_{\text{fin}}(\mathbb{C})$, $b \in \ell_1(\mathbb{C})$. □

6.1.2 Reduktion auf eine inhomogene Verfeinerungsgleichung

Es seien ein Skalenfaktor $s \in \mathbb{N}_{\ge 2}$ und eine Approximationsordnung A fixiert. Seien $a = sH_s(\mathcal{T})^A p$ und $\tilde{a} = sH_s(\mathcal{T})^A \tilde{p}$ zwei endliche Skalierungsfolgen, d.h. $p, \tilde{p} \in \ell_{\text{fin}}(\mathbb{C})$. Seien weiter $\varphi, \tilde{\varphi}$ zwei stetige Funktionen mit kompaktem Träger, die jeweils Lösungen der Verfeinerungsgleichungen $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ und $\tilde{\varphi} = \mathcal{D}_s \tilde{a}(\mathcal{T})\tilde{\varphi}$ sind. Diese seien ebenfalls zulässige Skalierungsfunktionen, d.h. insbesondere sollen $\sum_{n\in\mathbb{Z}} \varphi(x+n) = 1$ und $\sum_{n\in\mathbb{Z}} \tilde{\varphi}(x+n) = 1$ für alle $x \in \mathbb{R}$ erfüllt sein. Dann hat das Laurent–Polynom $\sum_{n\in\mathbb{Z}} \tilde{\varphi}(n) Z^n$ keine Nullstelle bei $Z = 1$, kann also modulo $(1-Z)^A$ invertiert werden. Aus dieser Invertierbarkeit folgt, dass es eine Folge $c \in \ell_{\text{fin}}(\mathbb{C})$ gibt, mit welcher

$$c(Z) \sum_{n\in\mathbb{Z}} \tilde{\varphi}(n) Z^n \equiv \sum_{k\in\mathbb{Z}} \varphi(k) Z^k \bmod (1-Z)^A \tag{6.1}$$

gilt. Die Verfeinerungsgleichung von φ (analog auch für $\tilde{\varphi}$) ergibt für die ganzzahlingen Punkte

$$\varphi(n) = \sum_{j\in\mathbb{Z}} a_j \varphi(sn-j) \quad \text{d.h.} \quad \{\varphi(n)\}_{n\in\mathbb{Z}} = \big(\downarrow s\big)\, a(\mathcal{T})\{\varphi(n)\}_{n\in\mathbb{Z}}\,.$$

Da die Folge a bzw. deren Laurent–Polynom $a(Z)$ das Haar–Polynom mit Vielfachheit A enthält, ergibt sich aus der Identität (6.1) und Satz 3.5.3, dass auch

$$\sum_{n\in\mathbb{Z}} \varphi(n) Z^{sn} \equiv H_s(Z)^A p(Z) \sum_{m\in\mathbb{Z}} \varphi(m) Z^m \bmod (1-Z)^A \tag{6.2}$$

gilt. Zusammen mit der vorhergehenden Gleichung (6.1) folgt daraus, wieder in Restklassen modulo $(1-Z)^A$ rechnend,

$$\begin{aligned} H_s(Z)^A p(Z) \sum_{m\in\mathbb{Z}} \varphi(m) Z^m \equiv \sum_{k\in\mathbb{Z}} \varphi(k) Z^{sk} &\equiv c(Z^s) \sum_{n\in\mathbb{Z}} \tilde{\varphi}(n) Z^{sn} \\ &\equiv c(Z^s) H_s(Z)^A \tilde{p}(Z) \sum_{n\in\mathbb{Z}} \tilde{\varphi}(n) Z^n \end{aligned} \tag{6.3}$$

Nochmals Gleichung (6.1) anwendend ergibt sich schließlich

$$p(Z)c(Z)\ H_s(Z)^A \sum_{n\in\mathbb{Z}} \tilde{\varphi}(n)Z^n \equiv c(Z^s)\tilde{p}(Z)\ H_s(Z)^A \sum_{n\in\mathbb{Z}} \tilde{\varphi}(n)Z^n \bmod (1-Z)^A\,. \tag{6.4}$$

Da nun die beiden letzten Faktoren keine Nullstelle in $Z = 1$ haben, sind sie modulo $(1-Z)^A$ invertierbar, können also aus dieser Identität gekürzt werden. Es muss somit eine Folge $\tilde{c} \in \ell_{\text{fin}}(\mathbb{C})$ geben, mit welcher

$$p(Z)c(Z) - c(Z^s)\tilde{p}(Z) = (1-Z)^A\tilde{c}(Z) \tag{6.5}$$

gilt.

Die Existenz der Folgen $c, \tilde{c}$ ist nicht von der Existenz der stetigen Skalierungsfunktionen abhängig, diese Folgen können allein aus den Folgen $p, \tilde{p}$ bestimmt werden.

Satz 6.1.3 *Seien $A \in \mathbb{N}_{\geq 1}$, $s \in \mathbb{N}_{>1}$ und $p, \tilde{p} \in \ell_{1,A}$ mit $\sum_{n\in\mathbb{Z}} p_n = \sum_{n\in\mathbb{Z}} \tilde{p}_n = 1$. Dann gibt es eine endliche Folge $c \in \ell_{\text{fin}}(\mathbb{C})$ mit $\sum_{n\in\mathbb{Z}} c_n = 1$ und eine Folge $\tilde{c} \in \ell_{\text{fin}}(\mathbb{C})$, so dass mit den Differenzenoperatoren zu p und $\tilde{p}$ gilt*

$$p(\mathcal{T})c - \tilde{p}(\mathcal{T})\left(\uparrow_{\mathrm{s}}\right)c = (1-\mathcal{T})^A\tilde{c}\,.$$

In jeder weiteren Lösung $(c^*, \tilde{c}^*)$ dieses Zerlegungsproblems unterscheidet sich c^* nur um ein Vielfaches von $(1-\mathcal{T})^A$ von c.

Beweis: Nach dem vorangegangenen Lemma 6.1.2 gibt es endliche Folgen $q, \tilde{q} \in \ell_{\text{fin}}(\mathbb{C})$ und „Restfolgen" $r, \tilde{r} \in \ell_1(\mathbb{C})$ mit $p = q + (1-\mathcal{T})^A r$ und $\tilde{p} = \tilde{q} + (1-\mathcal{T})^A \tilde{r}$. Jede Lösung $(c, \tilde{c})$ des Zerlegungsproblems mit $q, \tilde{q}$ anstelle von $p, \tilde{p}$ läßt sich einfach in eine Lösung des ursprünglichen Zerlegungsproblems umformen, denn aus

$$q(\mathcal{T})c - \tilde{q}(\mathcal{T})\left(\uparrow_{\mathrm{s}}\right)c = (1-\mathcal{T})^A\tilde{c}$$

folgt

$$p(\mathcal{T})c - \tilde{p}(\mathcal{T})\left(\uparrow_{\mathrm{s}}\right)c = (1-\mathcal{T})^A\left(\tilde{c} + r(\mathcal{T})c - \tilde{r}(\mathcal{T})\left(\uparrow_{\mathrm{c}}\right)\right)\,.$$

Das Problem in den endlichen Folgen läßt sich als Gleichung in Laurent–Polynomen schreiben,

$$q(Z)c(Z) - \tilde{q}(Z)c(Z^s) = (1-Z)^A\tilde{c}(Z).$$

Nach der Konstruktion in Lemma 6.1.2 kann vorausgesetzt werden, dass die Laurent–Polynome $q(Z), \tilde{q}(Z)$ schon „echte" Polynome sind. Durch die Variablensubstitution $Z = 1 + U$ erhalten wir ein einfach zu lösendes lineares Gleichungssystem. Seien $Q(U) := q(1+U)$, $\tilde{Q}(U) := \tilde{q}(1+U)$, $C(U) := c(1+U)$ und $\tilde{C}(U) := \tilde{c}(1+U)$, dann hat das Zerlegungsproblem die Form

$$Q(U)C(U) - \tilde{Q}(U)C(sU + \tbinom{s}{2}U^2 + \cdots + U^s) = U^A\tilde{C}(U)\,.$$

Für die Potenzen $1, U, \ldots, U^{A-1}$ ergibt sich durch Koeffizientenvergleich in der k–ten Potenz eine lineare homogene Gleichung in den Koeffizienten von C, welche nur die Koeffizienten $C_0, \ldots, C_k$ enthält. Dabei hat C_k in dieser Gleichung den Koeffizienten $1 - s^k$, denn es gilt

$Q(0) = \tilde{Q}(0) = 1$. Die Gleichung zu $k = 0$ ist trivial, womit $C_0 = 1$ gewählt werden kann. Nun können nacheinander die Koeffizienten $C_2, \dots, C_{A-1}$ bestimmt werden. Durch Einsetzen von $C(U)$ in die linke Seite ergeben sich auch die Koeffizienten von $\tilde{C}$. Durch Rücksubstitution erhalten wir eine Lösung als $c(Z) = C(Z-1)$ und $\tilde{c} = \tilde{C}(Z-1)$.

Die aus dem Koeffizientenvergleich gewonnenen Bedingungen für Lösungen $(c, \tilde{c})$ mit Träger $\operatorname{supp} c \subset [0, A-1] \cap \mathbb{Z}$ sind nicht nur hinreichend, sondern auch notwendig. Daher ist die oben konstruierte Lösung $(c_0, \tilde{c}_0)$ die einzige mit diesem Träger. Sei $(c^*, \tilde{c}^*) \in \ell_{\text{fin}}(\mathbb{C}) \times \ell_1(\mathbb{C})$, $c_1(1) = 1$, eine beliebige weitere Lösung des Zerlegungsproblems. Nach Lemma 6.1.2 gibt es dann eine endliche Folge d mit Träger in $\{0, \dots, A-1\}$ und eine endliche „Restfolge" r, so dass $c^* = d + (1-\mathcal{T})^A r$ gilt. Dann ist aber auch

$$p(\mathcal{T})d - \tilde{p}(\mathcal{T})\,(\uparrow \mathrm{s})\,d = (1-\mathcal{T})^A \left(\tilde{c}^* - p(\mathcal{T})r + s^A H(\mathcal{T})^A\,\tilde{p}(\mathcal{T})\,(\uparrow \mathrm{s})\,r\right)$$

und $d(1) = c^*(1) = 1$. D.h. d ist Teil einer Lösung des Zerlegungsproblems mit Träger in $\{0, \dots, A-1\}$ und daher identisch zu c, damit gilt $c^* = c + (1-\mathcal{T})^A r$. □

Die definierende Eigenschaft (6.5) der Folgen c und $\tilde{c}$ ausnutzend ergibt sich unter Multiplikation mit $sH_s(\mathcal{T})^A$ die Identität der Differenzenoperatoren

$$a(\mathcal{T})c(\mathcal{T}) - \tilde{a}(\mathcal{T})c(\mathcal{T}^s) = s^{1-A}(1-\mathcal{T}^s)^A\tilde{c}(\mathcal{T})\,.$$

Für die Differenz $\varepsilon := \varphi - c(\mathcal{T})\tilde{\varphi}$ der Skalierungsfunktionen gilt somit eine inhomogene Verfeinerungsgleichung

$$\begin{aligned}\varepsilon = \varphi - c(\mathcal{T})\tilde{\varphi} =& \mathcal{D}_s a(\mathcal{T})\varphi - \mathcal{D}_s c(\mathcal{T}^s)\tilde{a}(\mathcal{T})\tilde{\varphi}\\ =& \mathcal{D}_s a(\mathcal{T})\varphi - \mathcal{D}_s a(\mathcal{T})c(\mathcal{T})\tilde{\varphi} + s^{1-A}(1-\mathcal{T})^A\mathcal{D}_s\tilde{c}(\mathcal{T})\tilde{\varphi}\\ =& \mathcal{D}_s a(\mathcal{T})\varepsilon + s^{1-A}(1-\mathcal{T})^A\mathcal{D}_s\tilde{c}(\mathcal{T})\tilde{\varphi}\,.\end{aligned}$$

Da die Funktionen $\varphi, \tilde{\varphi}$ für die Motivation mit kompaktem Träger angenommen wurden, hat auch ε einen kompakten Träger, die Funktion $\eta : \mathbb{R} \to \mathbb{C}$ mit

$$x \mapsto \eta(x) := \sum_{n=0}^{\infty} \binom{A-1+n}{A-1} \varepsilon(x-n)$$

ist daher wohldefiniert und stetig. Für z innerhalb der Einheitskreisscheibe von $\mathbb{C}$ und ebenfalls im Sinne formaler Potenzreihen gilt mit der binomischen Reihe

$$(1-z)^{-A} = \sum_{n=0}^{\infty} \binom{-A}{n}(-z)^n = \sum_{n=0}^{\infty} \binom{n+A-1}{n} z^n = \sum_{n=0}^{\infty} \binom{A-1+n}{A-1} z^n\,.$$

Somit gilt auch $\varepsilon = (1-\mathcal{T})^A\eta$. Man überlegt sich leicht, dass die Folge der Funktionswerte $\{\eta(x+n)\}_{n\in\mathbb{Z}}$ für jede reelle Zahl der Form $x = \frac{m}{s^n}$ endlich sein muss, da dies für $x = 0$ und damit für $x \in \mathbb{Z}$ gilt, und mit der affinen Verfeinerungsgleichung für ε diese Eigenschaft von jedem $x \in \mathbb{R}$ auch auf $\frac{x}{s}$ vererbt wird. Aufgrund der Stetigkeit von η und dem kompakten

Träger von ε ergibt sich, dass auch η einen kompakten Träger hat. In der affinen Verfeinerungsgleichung, die nun die Form

$$(1-\mathcal{T})^A \eta = (1-\mathcal{T})^A s^{1-A} \mathcal{D}_s(p(\mathcal{T})\eta + \tilde{c}(\mathcal{T})\tilde{\varphi})$$

annimmt, kann daher der Faktor $(1-\mathcal{T})^A$ auf beiden Seiten gekürzt werden.

Ist also die stetige Skalierungsfunktion $\tilde{\varphi}$ vorgegeben, so kann φ aus einer Lösung der reduzierten affinen Verfeinerungsgleichung

$$\eta = s^{1-A} \mathcal{D}_s(p(\mathcal{T})\eta + \tilde{c}(\mathcal{T})\tilde{\varphi}) \tag{6.6}$$

als

$$\varphi = c(\mathcal{T})\tilde{\varphi} + (1-\mathcal{T})^A \eta$$

bestimmt werden. Die Existenz von η und damit der Skalierungsfunktion φ ist gesichert, wenn der Operator $\mathcal{D}_s p(\mathcal{T})$ in einem zu wählenden Banachraum kontraktiv ist. Sind η und somit φ stetig, so sind sie bereits durch die Wertefolgen $\{\eta(n)\}_{n\in\mathbb{Z}}$ bzw. $\{\varphi(n)\}_{n\in\mathbb{Z}}$ eindeutig bestimmt.

6.1.3 Generisches Existenztheorem

Wir wollen verschiedene Normen betrachten, welche auf dem Raum $C_c(\mathbb{R})$ der stetigen Funktionen mit kompaktem Träger definiert sind. Für die nachfolgend dargestellte Existenztheorie der Verfeinerungsgleichung müssen diese Normen einige Bedingungen erfüllen.

Sei $\|\cdot\| : C_c(\mathbb{R}) \to \mathbb{R}_+$ eine Norm und $(B, \|\cdot\|)$ die Vervollständigung des normierten Raumes $(\mathbb{C}_c(\mathbb{R}), \|\cdot\|)$. Dann soll gelten:

- Die Fortsetzungen der Translationsoperatoren $\mathcal{T}^k$ sind für alle $k \in \mathbb{Z}$ Isometrien von B. Die Fortsetzungen der Dilatationsoperatoren $\mathcal{D}_s$ sind für alle $s \in \mathbb{N}_{>1}$ beschränkt und invertierbar.
- Gilt $f - \mathcal{T}f = g$ mit $f, g \in B$ und hat g einen kompakten Träger, dann hat f ebenfalls einen kompakten Träger und ist eindeutig durch g bestimmt.

Lemma 6.1.4 *Sei $(B, \|\cdot\|)$ ein Banachraum, dessen Norm den obigen Bedingungen genügt. Seien $A \in \mathbb{N}$, $f \in B$, $a \in \ell_{\text{fin}}(\mathbb{C})$ und gelte*

$$(1-\mathcal{T})f = a(\mathcal{T})\beta_{A-1}\,.$$

Dann gibt es ein $b \in \ell_{\text{fin}}(\mathbb{C})$, so dass $f = b(\mathcal{T})\beta_{A-1}$ und $a(\mathcal{T}) = (1-\mathcal{T})b(\mathcal{T})$ gelten.

Beweis: Nach der zweiten Forderung an den Banachraum B hat f einen kompakten Träger. Seien $M, N \in \mathbb{N}$ so groß, dass

$$\operatorname{supp} a \subset [-N, N], \quad N + A < M \quad \text{und} \quad \operatorname{supp} f \subset [1-M, M-1]$$

gelten. Dann gilt $f(M) - f(-M) = 0$, und mit

$$\operatorname{supp} \beta_{A-1} = [0, A] \quad \text{sowie} \quad \sum_{m=0}^{A} \beta_{A-1}(m) = 1$$

folgt

$$0 = \Big((1-\mathcal{T}^{2M})f\Big)(M) = \sum_{k=0}^{2M-1}\Big(\mathcal{T}^k a(\mathcal{T})\beta_{A-1}\Big)(M) = \sum_{k=0}^{2M-1}\sum_{n=-N}^{N} a_n\Big(\mathcal{T}^{k+n}\beta_{A-1}\Big)(M)$$
$$= \sum_{n=-N}^{N} a_n \sum_{m=-M-1-n}^{M-n} \beta_{A-1}(m) = \sum_{n=-N}^{N} a_n\,.$$

Damit hat das Laurent–Polynom $a(Z) = \sum_{n=-N}^{N} a_n Z^n$ eine Nullstelle bei $Z = 1$, es kann also der Linearfaktor $(1-Z)$ abgespalten werden. Sei $b \in \ell_{\text{fin}}(\mathbb{C})$ die Koeffizientenfolge des Quotienten, $a(Z) = (1-Z)b(Z)$. Daher gilt

$$0 = (1-\mathcal{T})(f - b(\mathcal{T})\beta_{A-1})\,,$$

und wegen der eindeutigen Lösbarkeit dieser Gleichung in B muss $f = b(\mathcal{T})\beta_{A-1}$ gelten. □

Eine einfache Folgerung ist, dass auch aus den Haar–Polynomen gebildete Differenzenoperatoren gekürzt werden können. Denn ist $H_s(\mathcal{T})f = H_s(\mathcal{T})g$, so folgt $(1-\mathcal{T}^s)(f-g) = 0$. Damit muss $\mathcal{D}_s(f-g) = 0$ gelten, also auch $f = g$.

Satz 6.1.5 *Sei $(B, \|\cdot\|)$ ein Banachraum, der $(\mathbb{C}_c(\mathbb{R}), \|\cdot\|)$ als dichte Teilmenge enthält und dessen Norm die oben gestellten Bedingungen erfüllt. Seien $s \in \mathbb{N}_{>1}$ und $A \in \mathbb{N}_{>0}$ fixiert und sei $p \in \ell_{1,A}(\mathbb{Z})$ eine Folge mit $\sum_{n\in\mathbb{Z}} p_n = 1$, für die der Operator $s^{1-A}\mathcal{D}_s p(\mathcal{T}) : B \to B$ kontraktiv ist.*

Dann gibt es eine endliche Folge $c \in \ell_{\text{fin}}(\mathbb{C})$ mit $\sum_{n\in\mathbb{Z}} c_n = 1$ und ein $\eta \in B$, so dass $\varphi := c(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A\eta$ eine Lösung der Verfeinerungsgleichung

$$\varphi = \mathcal{D}_s a(\mathcal{T})\varphi, \qquad a(\mathcal{T}) = sH_s(\mathcal{T})^A p(\mathcal{T})$$

ist. Jede weitere Lösung der Verfeinerungsgleichung dieser Form stimmt mit φ überein.

Beweis: Nach Satz 6.1.3 gibt es Folgen $c \in \ell_{\text{fin}}(\mathbb{C})$ und $\tilde{c} \in \ell_1(\mathbb{C})$ mit

$$p(\mathcal{T})c(\mathcal{T}) - c(\mathcal{T}^s) = (1-\mathcal{T})^A\tilde{c}(\mathcal{T})\,. \tag{6.7}$$

Nach Voraussetzung ist $s^{1-A}\mathcal{D}_s p(\mathcal{T})$ kontraktiv, somit hat die Fixpunktgleichung

$$\eta = s^{1-A}\mathcal{D}_s\Big(\tilde{c}(\mathcal{T})\beta_{A-1} + p(\mathcal{T})\eta\Big) \tag{6.8}$$

genau eine Lösung $\eta \in B$. Die Fixpunktgleichung kann nun – unter Beachtung von $(1-\mathcal{T})\mathcal{D}_s = \mathcal{D}_s(1-\mathcal{T}^s) = s\mathcal{D}_s H(\mathcal{T})(1-\mathcal{T})$ – umgeformt werden zu

$$\begin{aligned}(1-\mathcal{T})^A\eta =& s\mathcal{D}_s H_s(\mathcal{T})^A(1-\mathcal{T})^A\big(\tilde{c}(\mathcal{T})\beta_{A-1} + p(\mathcal{T})\eta\big)\\ =& s\mathcal{D}_s(1-\mathcal{T})^A\tilde{c}(\mathcal{T})H_s(\mathcal{T})^A\beta_{A-1} + s\mathcal{D}_s H_s(\mathcal{T})^A p(\mathcal{T})(1-\mathcal{T})^A\eta\,.\end{aligned}$$

Unter Benutzung von Gleichung (6.7) und der Verfeinerungsgleichung der B–Splines erhalten wir daraus

$$\begin{aligned}(1-\mathcal{T})^A\eta =& s\mathcal{D}_s\Big(p(\mathcal{T})c(\mathcal{T}) - c(\mathcal{T}^s)\Big)H_s(\mathcal{T})^A\beta_{A-1} + \mathcal{D}_s a(\mathcal{T})(1-\mathcal{T})^A\eta\\ =& \mathcal{D}_s a(\mathcal{T})\Big(c(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A\eta\Big) - c(\mathcal{T})\beta_{A-1}\,.\end{aligned}$$

Die Funktion $\varphi := c(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A\eta \in B$ ist somit Lösung der Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$.

Seien $c_1 \in \ell_{\text{fin}}(\mathbb{C})$ und $\eta_1 \in B$ derart, dass Bestandteile einer weiteren Lösung

$$\sum_{n\in\mathbb{Z}} c_{1,n} = 1 \quad \text{und} \quad \varphi_1 = c_1(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A\eta_1$$

eine weitere Lösung der Verfeinerungsgleichung ist. Unter Umkehrung der eben vorgenommenen Rechnung erhalten wir aus den Verfeinerungsgleichungen von φ_1 und β_{A-1} die Beziehung

$$\begin{aligned}(1-\mathcal{T})^A\left(\eta_1 - s^{1-A}\mathcal{D}_s p(\mathcal{T})\eta_1\right) =& (1-\mathcal{T})^A\eta_1 - \mathcal{D}_s a(\mathcal{T})(1-\mathcal{T})^A\eta_1\\ =& \varphi - c_1(\mathcal{T})\beta_{A-1} - \mathcal{D}_s a(\mathcal{T})\left(\varphi_1 - c_1(\mathcal{T})\beta_{A-1}\right)\\ =& s\mathcal{D}_s H_s(\mathcal{T})^A\left(p(\mathcal{T})c_1(\mathcal{T}) - c_1(\mathcal{T}^s)\right)\beta_{A-1}\,.\end{aligned}$$

Nach Lemma 6.1.2 gibt es nun eine endliche Folge $q \in \ell_{\text{fin}}(\mathbb{C})$ und eine Restfolge $r \in \ell_1(\mathbb{C})$ mit $p = q + (1-\mathcal{T})^A r$. Wir können die eben gefundene Beziehung somit auf die Form von Lemma 6.1.4 bringen. Seien dazu $f := s^{A-1}\mathcal{D}_s^{-1}\eta_1 - p(\mathcal{T})\eta_1 \in B$ und $b := q(\mathcal{T})c_1 - \left(\uparrow\mathrm{s}\right)c_1 \in \ell_{\text{fin}}(\mathbb{C})$. Dann gilt

$$\begin{aligned}(1-\mathcal{T})^A s^{1-A}\mathcal{D}_s f =& s\mathcal{D}_s H_s(\mathcal{T})^A\left(b(\mathcal{T}) + (1-\mathcal{T})^A r(\mathcal{T})c(\mathcal{T})\right)\beta_{A-1}\\ \iff (1-\mathcal{T})^A\left(f - r(\mathcal{T})c(\mathcal{T})\beta_{A-1}\right) =& b(\mathcal{T})\beta_{A-1}\end{aligned}$$

Nach Lemma 6.1.4 muss es eine endliche Folge $\tilde{c}_1 \in \ell_{\text{fin}}(\mathbb{C})$ geben, so dass

$$(1-\mathcal{T})^A\,\tilde{c}_1 = b = q(\mathcal{T})c_1 - \left(\uparrow\mathrm{s}\right)c_1 = p(\mathcal{T})c_1 - \left(\uparrow\mathrm{s}\right)c_1 - (1-\mathcal{T})^A\,r(\mathcal{T})c_1$$

gilt. Somit ist $(c_1, \tilde{c}_1 + r(\mathcal{T})c_1)$ eine Lösung des Zerlegungsproblems von Satz 6.1.3 zur Folge $p \in \ell_{1,A}(\mathbb{C})$, es gibt daher eine endliche Folge $d \in \ell_{\text{fin}}(\mathbb{C})$ mit $c_1 = c + (1-\mathcal{T})^A\,d$. Folglich gilt

$$\varphi_1 = c_1\beta_{A-1} + (1-\mathcal{T})^A\,\eta_1 = c\beta_{A-1} + (1-\mathcal{T})^A(\eta_1 + d(\mathcal{T})\beta_{A-1})\,.$$

Damit ist $\eta_1 + d(\mathcal{T})\beta_{A-1}$ eine Lösung der Fixpunktgleichung (6.8), muss also mit η übereinstimmen, es gilt $\varphi_1 = \varphi$. □

Es ist oft sinnvoll, eine gegebene Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ nicht nur auf der Skala $s \in \mathbb{N}_{>1}$ zu betrachten, sondern auch Iterationen davon auf Skalen s^N, $N \in \mathbb{N}$. Denn hat die Verfeinerungsgleichung eine Lösung $\varphi \in B$, so erfüllt diese auch die iterierte Verfeinerungsgleichung

$$\varphi = (\mathcal{D}_s a(\mathcal{T}))^N\varphi = \mathcal{D}_{s^N} a(\mathcal{T})a(\mathcal{T}^s)\ldots a(\mathcal{T}^{s^{N-1}})\varphi\,.$$

Gilt $a = sH_s(\mathcal{T})^A p(\mathcal{T})$ mit $A \in \mathbb{N}_{>0}$ und $p \in \ell_{1,A}(\mathbb{C})$, so gibt es eine Lösung der Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$, wenn der Operator $s^{1-A}\mathcal{D}_s p(\mathcal{T})$ kontraktiv ist. Seien Folgen

a_N, p_N durch $a_N(\mathcal{T}) := a(\mathcal{T})a(\mathcal{T}^s)\dots a(\mathcal{T}^{s^{N-1}})$ und $p_N(\mathcal{T}) := p(\mathcal{T})p(\mathcal{T}^s)\dots p(\mathcal{T}^{s^{N-1}})$ gegeben, dann gelten

$$a_N = s^N H_{s^N}(\mathcal{T})^A p_N \quad \text{und} \quad \mathcal{D}_{s^N} p_N(\mathcal{T}) = (\mathcal{D}_s p(\mathcal{T}))^N .$$

Daher ist auch der affin lineare Fixpunktoperator der iterierten Verfeinerungsgleichung kontraktiv.

Ist umgekehrt der affin lineare Fixpunktoperator der iterierten Verfeinerungsgleichung kontraktiv, so gibt es eine Lösung $\tilde{\varphi}$ der Verfeinerungsgleichung $\tilde{\varphi} = (\mathcal{D}_s a(\mathcal{T}))^N \tilde{\varphi}$. Dann ergibt sich aus dieser durch

$$\varphi := \frac{1}{N} \sum_{k=0}^{N-1} (\mathcal{D}_s a(\mathcal{T}))^k \, \tilde{\varphi}$$

auch eine Lösung der gegebenen Verfeinerungsgleichung.

Jedoch können die Abschätzungen der Operatornorm für den iterierten Operator $s^{N(1-A)}(\mathcal{D}_s p(\mathcal{T}))^N$ kleiner ausfallen als die Potenz der Abschätzungen des einfachen Operators. Insbesondere ist es möglich, dass die Kontraktivität nur für eine der iterierten Verfeinerungsgleichungen nachgewiesen werden kann, und damit auch die Existenz einer Lösung der einfachen Verfeinerungsgleichung, obwohl die gefundenen Abschätzungen für diese keine direkte Aussage über die Existenz einer Lösung erlauben.

6.2 Analytische Eigenschaften von Skalierungsfunktionen

Wir werden nun die Voraussetzungen des Existenzsatzes für die L^q–Normen und gemischte Normen nachprüfen und erhalten daraus verschiedene Kriterien an die Folge $p \in \ell_{1,A}(\mathbb{C})$, welche Eigenschaften der Lösungen φ der Verfeinerungsgleichung zur Skalierungsfolge $a = sH_s(\mathcal{T})^A p$ ergeben. Insbesondere kann aus diesen Bedingungen auch auf Stetigkeit und Differenzierbarkeit der Skalierungsfunktion φ geschlossen werden.

6.2.1 Existenz von Lösungen in $L^q(\mathbb{R})$

Sei $1 \le q < \infty$. Dann enthält der Funktionenraum $B = L^q(\mathbb{R})$ die Menge $C_c(\mathbb{R})$ der stetigen Funktionen mit kompaktem Träger als dichte Teilmenge. Es sind die Zulässigkeitsbedingungen an $L^q(\mathbb{R})$ zu prüfen.

Die Norm auf $L^q(\mathbb{R})$ ist translationsinvariant. Für beliebige Stauchungen und Streckungen $\mathcal{D}_s$, $s > 0$ gilt

$$\|\mathcal{D}_s f\|_{L^q} = \sqrt[q]{s^{-1}} \|f\|_{L^q} .$$

Lemma 6.2.1 *Es seien $f, g \in L^q(\mathbb{R})$, $q \in [1, \infty)$. Weiter habe g kompakten Träger und es gelte $(1 - \mathcal{T})f = g$. Dann hat f ebenfalls einen kompakten Träger. Ist $g = 0$, so folgt $f = 0$.*

Beweis: Es gibt ein $M \in \mathbb{N}$ mit $\operatorname{supp} g \subset [-M, M]$. Die Reihe

$$\tilde{f} := \sum_{k=0}^{\infty} \mathcal{T}^k g$$

hat über jedem Intervall $[m, m+1]$, $m \in \mathbb{Z}$, nur endlich viele nichtverschwindende Summanden, die Einschränkung von $\tilde{f}$ auf dieses Intervall ist ein Element von $L^q([m, m+1])$. $\tilde{f}$ verschwindet auf dem Intervall $(-\infty, -M)$ nach Konstruktion.

Über jedem beschränkten Intervall kann nun die Wirkung des Differenzenoperators betrachtet werden, es gilt

$$(1-\mathcal{T})\tilde{f} = \sum_{k=0}^{\infty} \mathcal{T}^k g - \sum_{k=1}^{\infty} \mathcal{T}^k g = g\,.$$

Damit gilt $(1-\mathcal{T})(f-\tilde{f}) = 0$, die Differenz $(f-\tilde{f})$ ist also periodisch mit Periode 1. Auf Intervallen $[m, m+1]$ mit $m < -M$ verschwindet jedoch $\tilde{f}$, die Einschränkungen von f auf ein solches Intervall werden mit gegen $-\infty$ fallendem m beliebig klein. Die einzige periodische Funktion mit diesen Eigenschaften ist die Nullfunktion. Somit gilt $f = \tilde{f} = \sum_{k=0}^{\infty} \mathcal{T}^k g$.

Der Träger von $\mathcal{T}^k g$ ist der um k verschobene Träger von g, d.h. in $[-M+k, M+k]$ enthalten. Dieses Intervall hat einen nichttrivialen Durchschnitt mit dem Intervall $[m, m+1]$, wenn $k + M > m$ und $k - M < m + 1$ gelten, d.h. für

$$m - M + 1 \le k \le m + M\,.$$

Für $m \ge M - 1$ ist daher die Einschränkung von f auf das Intervall $[m, m+1]$ durch die endliche Summe

$$f = \sum_{k=m-M+1}^{m+M} \mathcal{T}^k g = \mathcal{T}^m \sum_{k=-M+1}^{M} \mathcal{T}^k g$$

gegeben, d.h. f ist auf dem Intervall $[M-1, \infty)$ periodisch mit Periode 1. Da andererseits $f \in L^q(\mathbb{R})$ gilt, muss f auch auf $[M-1, \infty)$ verschwinden, es gilt somit $\operatorname{supp} f \subset [-M, M-1]$ und $\sum_{k=-M+1}^{M} \mathcal{T}^k g = 0$.

Ist $g = 0$, so ist $f \in L^q(\mathbb{R})$ periodisch, was nur für $f = 0$ möglich ist. □

Satz 6.2.2 *Seien $q \in [1, \infty)$, $s \in \mathbb{N}_{>1}$ und $A \in \mathbb{N}_{>0}$ fixiert.*

Sei $p \in \ell_{1,A}(\mathbb{Z})$ eine Folge mit $\sum_{n\in\mathbb{Z}} p_n = 1$ und

$$\|p\|_{\ell_1} = \sum_{n\in\mathbb{Z}} |p_n| < s^{A+p^{-1}-1}\,.$$

Ferner sei $a(\mathcal{T}) = sH_s(\mathcal{T})^A p(\mathcal{T})$. Dann gibt es genau eine Funktion $\varphi \in L^q(\mathbb{R})$, die die Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ löst und als $\varphi = c(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A \eta$ mit $c \in \ell_{\text{fin}}(\mathbb{C})$, $\sum_{n\in\mathbb{Z}} c_n = 1$ und $\eta \in L^q(\mathbb{R})$ darstellbar ist.

Beweis: Dieser Satz ist eine Folge des generischen Existenzsatzes 6.1.5. Die Bedingungen an die Norm von $L^q(\mathbb{R})$, $1 \le q < \infty$, wurden bereits geprüft, es verbleibt die Kontraktivität von $s^{1-A}\mathcal{D}_s p(\mathcal{T})$ zu zeigen. Für eine beliebige Funktion $f \in L^p(\mathbb{R})$ gilt

$$\|s^{1-A}\mathcal{D}_s p(\mathcal{T}) f\|_{L^q} \le s^{1-A}\sqrt[q]{s^{-1}}\|p(\mathcal{T})f\|_{L^q} \le s^{1-A-q^{-1}}\|p\|_1\,\|f\|_{L^q}\,.$$

Der konstante Faktor $\left(s^{1-A-q^{-1}}\|p\|_1\right)$ auf der rechten Seite ist nach Voraussetzung kleiner als 1, womit auch die Kontraktivität gezeigt ist. □

6.2.2 Existenz von stetigen Lösungen

Seien $p \in \ell_{\text{fin}}(\mathbb{R})$ eine beliebige endliche Folge und $f \in C_c(\mathbb{R})$ eine stetige Funktion mit kompaktem Träger. Dann ist $g := \mathcal{D}_s p(\mathcal{T}) f$ ebenfalls eine stetige Funktion mit kompaktem Träger, da die zugehörige Funktionenreihe $\sum_{n\in\mathbb{Z}} p_n(\mathcal{T}^n f)$ nur endlich viele Summanden hat. Die Werte von g sind somit wohldefiniert und ergeben sich als

$$g(x) = \sum_{n\in\mathbb{Z}} p_n f(sx - n) \,.$$

In $g(x)$ gehen also die Werte der Folge $\{f(sx+n)\}_{n\in\mathbb{Z}}$ ein. Die Glieder der entsprechenden Wertefolge von g sind

$$g(x+m) = \sum_{n\in\mathbb{Z}} p_n f(sx + sm - n) = \sum_{n\in\mathbb{Z}} p_{n+sm} f(sx - n) \,.$$

Bezeichnen wir der Übersicht halber die Folgen der Funktionswerte mit

$$\vec{f}(x) := \{f(x+n)\}_{n\in\mathbb{Z}} \text{ und } \vec{g}(x) := \{g(x+n)\}_{n\in\mathbb{Z}} \,,$$

so kann obige Beziehung kurz als $\vec{g}(x) = \left(\downarrow \mathrm{s}\right) p(\mathcal{T})\vec{f}(sx)$ geschrieben werden. Seien $(p_{(0)}, p_{(1)}, \ldots, p_{(s-1)}) := \left(\downdownarrows \mathrm{s}\right) p$ die Polyphasenteilfolgen von p, dann gilt für deren Differenzenoperatoren

$$p(\mathcal{T}) = p_{(0)}(\mathcal{T}^s) + p_{(1)}(\mathcal{T}^s)\mathcal{T} + \cdots + p_{(s-1)}\mathcal{T}^{s-1}$$

und damit

$$\vec{g}(x) = \sum_{k=0}^{s-1} p_{(k)}(\mathcal{T}) \left(\downarrow \mathrm{s}\right)\left(\mathcal{T}^k \vec{f}(sx)\right) \,.$$

Sei ein $q \in [1,\infty]$ fixiert. Da f einen kompakten Träger hat, ist $\vec{f}(x)$ für jedes $x \in \mathbb{R}$ eine endliche Folge und damit in $\ell_q(\mathbb{C})$ enthalten. Nach Konstruktion gilt dies auch für $\vec{g}(x)$, und es gilt

$$\|\vec{g}(x)\|_{\ell_q} \le \sum_{k=0}^{s-1} \|p_{(k)}\|_{\ell_1} \left\| \left(\downarrow \mathrm{s}\right)\left(\mathcal{T}^k \vec{f}(sx)\right)\right\|_{\ell_q} \,.$$

Die rechte Seite hat die Form eines Skalarproduktes und kann mittels der Hölder–Ungleichung weiter umgeformt werden. Sei dazu $\tilde{q} \in [1,\infty]$ so bestimmt, dass $q^{-1} + \tilde{q}^{-1} = 1$ gilt. Bei $q = 1$ sei $\tilde{q} := \infty$, bei $q = \infty$ sei $\tilde{q} := 1$. Dann gilt in den q– und $\tilde{q}$–Normen des $\mathbb{R}^s$ für Vektoren $\mathbf{v} = (v_0, \ldots, v_{s-1})$ und $\mathbf{w} = (w_0, \ldots, w_{s-1})$

$$\sum_{k=0}^{s-1} v_k w_k \le \|\mathbf{v}\|_{\tilde{q}} \, \|\mathbf{w}\|_q = \left(|v_0|^{\tilde{q}} + \ldots |v_{s-1}|^{\tilde{q}}\right)^{(\tilde{q}^{-1})} \left(|w_0|^q + \ldots |w_{s-1}|^q\right)^{(q^{-1})} \,.$$

Für den Vektor $\mathbf{w}$ mit $w_k = \| \left(\downarrow \mathrm{s}\right) \left(\mathcal{T}^k \vec{f}(sx)\right) \|_{\ell_q}$ erhalten wir für die q–Norm des $\mathbb{R}^s$

$$\begin{aligned}\left(\|\mathbf{w}\|_q\right)^q &= \sum_{k=0}^{s-1} \left\| \left(\downarrow \mathrm{s}\right) \left(\mathcal{T}^k \vec{f}(sx)\right) \right\|_{\ell_q}^q = \sum_{k=0}^{s-1} \sum_{n\in\mathbb{Z}} |f(sx-k+sn)|^q \\ &= \sum_{n\in\mathbb{Z}} |f(sx+n)|^q = \left\| \vec{f}(sx) \right\|_{\ell_q}^q .\end{aligned}$$

Sei der Vektor $\mathbf{v} \in \mathbb{R}^s$ komponentenweise durch die Normen der Polyphasen von p gegeben, d.h. $v_k := \|p_{(k)}\|_{\ell_1}, k = 0, \ldots, s-1$. Wir bezeichnen die Norm von $\mathbf{v}$ mit $\left\| \left(\downdownarrows \mathrm{s}\right) p \right\|_{(1,\tilde{q})} := \|\mathbf{v}\|_{\tilde{q}}$. Zusammenfassend erhalten wir

$$\|\vec{g}(x)\|_{\ell_q} \leq \sum_{k=0}^{s-1} v_k w_k \leq \|\mathbf{v}\|_{\tilde{q}} \|\mathbf{w}\|_q = \| \left(\downdownarrows \mathrm{s}\right) p\|_{(1,\tilde{q})} \left\| \vec{f}(sx) \right\|_{\ell_q} . \tag{6.9}$$

Wir können nun gemischte Normen auf $C_c(\mathbb{R})$ definieren, indem wir die stetige Funktion $h : [0,1] \to \mathbb{R}$, die durch ℓ_q–Norm der Folgen $\vec{f}(\cdot)$ gegeben ist, $h(x) := \left\|\vec{f}(x)\right\|_{\ell_q}$, als Element von $L^r([0,1])$ auffassen, $r \in [1,\infty]$.

Definition 6.2.3 (vgl. [BZ97] und dort angegebene Quellen)
Sei $\mathbb{T} := [0,1)$. Mit $L^{q,r}(\mathbb{Z}\times\mathbb{T}) = L^{q,r}(\mathbb{Z}\times\mathbb{T},\mathbb{C})$ bezeichnen wir die Vervollständigung von $\left(C_c(\mathbb{R}), \|\cdot\|_{(q,r)}\right)$, wobei die Norm für alle $f \in C_c(\mathbb{R})$ durch

$$\|f\|_{(p,r)} := \left\| \|\vec{f}(\cdot)\|_{\ell_q} \right\|_{L^r([0,1])} = \left(\int_0^1 \left(\sum_{n\in\mathbb{Z}} |f(x+n)|^q \right)^{\frac{r}{q}} \mathrm{d}x \right)^{\frac{1}{r}}$$

definiert ist.

Nach Konstruktion ist $\|\cdot\|_{(q,r)}$ zunächst eine Halbnorm. Um zu prüfen, ob dies auch eine Norm auf $C_c(\mathbb{R})$ ist, verbleibt zu zeigen, dass $\|f\|_{(p,r)} = 0$ nur für $f = 0$ gilt. Da $\|\vec{f}(\cdot)\|_{\ell_q}$ über jedem Intervall eine endliche Summe ist, ist dies eine stetige periodische Funktion. Gilt $\|f\|_{(p,r)} = 0$, so müssen die Folgennormen überall verschwinden, damit aber auch alle Funktionswerte von f, denn jedes $x \in \mathbb{R}$ kann als $x = y + n$ mit $n \in \mathbb{Z}$ und $y \in [0,1]$ dargestellt werden, aus $\|\vec{f}(y)\|_{\ell_q} = 0$ folgt $f(x) = 0$.

Lemma 6.2.4 *Seien $1 \leq q \leq q' \leq \infty$ und $1 \leq r' \leq r \leq \infty$. Dann gilt für beliebige $f \in C_c(\mathbb{R})$*

$$\|f\|_{(q',r')} \leq \|f\|_{(q,r)} . \tag{6.10}$$

Insbesondere gelten

- *$L^{1,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^{q,r}(\mathbb{Z}\times\mathbb{T}) \subset L^q(\mathbb{R}) \cap L^r(\mathbb{R})$ für beliebige $1 \leq q \leq r < \infty$ und*
- *$L^{1,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^{q,\infty}(\mathbb{Z}\times\mathbb{T}) \leq L^q(\mathbb{R}) \cap C_b(\mathbb{R})$ für beliebiges $1 \leq q \leq \infty$, wobei $C_b(\mathbb{R})$ der Raum der beschränkten stetigen Funktionen ist.*

Beweis: Sei $\alpha := \frac{q'}{q} \geq 1$. Für jede Folge $c \in \ell_1(\mathbb{C})$ folgt aus der Dreiecksungleichung die Ungleichung $\|c\|_{\ell_\alpha} \leq \|c\|_{\ell_1}$. Damit gilt für jede Folge $c \in \ell_{\mathrm{fin}}(\mathbb{C})$

$$\|c\|_{\ell_{q'}}^q = \|\{|c_n|^q\}_{n\in\mathbb{Z}}\|_{\ell_\alpha} \leq \|\{|c_n|^q\}_{n\in\mathbb{Z}}\|_{\ell_1} = \|c\|_{\ell_q}^q .$$

Daher gilt diese Ungleichung auch für jedes $c \in \ell_q(\mathbb{C})$ und es folgt $\|f\|_{(q',r)} \leq \|f\|_{(q,r)}$ für jedes $f \in C_c(\mathbb{R})$.

Sei nun $\alpha := \frac{r}{r'} \geq 1$. Nach der Hölder–Ungleichung gilt für jedes $h \in L^\alpha([0,1])$ die Ungleichung $\|h\|_{L^1} \leq \|h\|_{L^\alpha}$. Somit gilt für jede stetige Funktion $h \in C([0,1])$

$$\|h\|^{r'}_{L^{r'}([0,1])} = \left\| \, |h(\cdot)|^{r'} \, \right\|_{L^1([0,1])} \leq \left\| \, |h(\cdot)|^{r'} \, \right\|_{L^\alpha([0,1])} = \|h\|^{r'}_{L^r([0,1])} \, .$$

Da $h := \|\vec{f}(\cdot)\|_{\ell_{q'}}$ für jedes $f \in C_c(\mathbb{R})$ stetig ist, folgt $\|f\|_{(q',r')} \leq \|f\|_{(q',r)}$. Insgesamt folgt also die Ungleichung (6.10) $\|f\|_{(q',r')} \leq \|f\|_{(q,r)}$.

Nun gilt für $f \in C_c(\mathbb{R})$ und $1 \leq r < \infty$

$$\|f\|^r_{(r,r)} = \int_0^1 \sum_{n\in\mathbb{Z}} |f(n+x)|^r \, \mathrm{d}x = \int_{\mathbb{R}} |f(x)|^r \, \mathrm{d}x = \|f\|^r_{L^r} \, .$$

Da $C_c(\mathbb{R})$ eine dichte Teilmenge von $L^r(\mathbb{R})$ ist, folgt $L^{r,r}(\mathbb{Z}\times\mathbb{T}) = L^r(\mathbb{R})$. Für $q \leq r$ gelten die Ungleichungen

$$\|f\|_{L^r} = \|f\|_{(r,r)} \leq \|f\|_{(q,r)} \text{ und } \|f\|_{L^q} = \|f\|_{(q,q)} \leq \|f\|_{(q,r)} \, .$$

Es sind jeweils beide Seiten der Ungleichung stetig in $f \in C_c(\mathbb{R})$, womit sie auch in der Vervollständigung $L^{q,r}(\mathbb{Z}\times\mathbb{T})$ gültig sind. Somit gilt

$$L^{q,r}(\mathbb{Z}\times\mathbb{T}) \subset L^q(\mathbb{R}) \cap L^r(\mathbb{R}) \, .$$

Analog folgt aus Ungleichung (6.10) auch

$$C_c(\mathbb{R}) \subset L^{1,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^{q,r}(\mathbb{Z}\times\mathbb{T}) \, .$$

Ist $r = \infty$, so ist für jedes $f \in C_c(\mathbb{R})$ die Gleichung

$$\|f\|_{(\infty,\infty)} = \sup_{x\in[0,1]} \sup_{n\in\mathbb{Z}} |f(x+n)| = \sup_{y\in\mathbb{R}} |f(y)| = \|f\|_{L^\infty}$$

erfüllt und damit $L^{q,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^\infty(\mathbb{R})$.

Weiter folgt aus $\sup_{x\in\mathbb{R}} |f(x)| \leq \|f\|_{(q,\infty)}$ für $f \in C_c(\mathbb{R})$, dass jede Cauchyfolge aus $\left(C_c(\mathbb{R}), \|\cdot\|_{(q,\infty)}\right)$ punktweise und gleichmäßig konvergiert, also eine stetige, beschränkte Funktion als Grenzwert hat. Somit gilt auch

$$L^{1,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^{q,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^q(\mathbb{R}) \cap C_b(\mathbb{R})$$

□

Lemma 6.2.5 *Seien $q, r \in [1,\infty]$, $s \in \mathbb{N}_{>1}$ und $p \in \ell_1(\mathbb{C})$ beliebig. Dann ist die Zahl $\| \left(\downarrow\downarrow s\right) p\|_{(1,\tilde{q})}$ mit $q^{-1} + \tilde{q}^{-1} = 1$ eine Schranke des Operators*

$$\mathcal{D}_s p(\mathcal{T}) : L^{q,r}(\mathbb{Z}\times\mathbb{T}) \to L^{q,r}(\mathbb{Z}\times\mathbb{T})$$

Beweis: Ist $p \in \ell_{\text{fin}}(\mathbb{C})$ wie oben eine endliche Folge und $f \in C_c(\mathbb{R})$, so erhalten wir für $r < \infty$ aus Ungleichung 6.9 die Abschätzung der Norm von $g = \mathcal{D}_s p(\mathcal{T}) f$ zu

$$\begin{aligned}\|g\|_{(q,r)}^r = \int_0^1 \|\vec{g}(x)\|_{\ell_q}^r \, \mathrm{d}x &\leq \| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})}^r \int_0^s \|\vec{f}(y)\|_{\ell_q}^r \tfrac{1}{s} \, \mathrm{d}y \\ &= \| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})}^r \int_0^1 \|\vec{f}(y)\|_{\ell_q}^r \, \mathrm{d}y = \| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})}^r \|f\|_{(p,r)}^r \,, \end{aligned} \tag{6.11}$$

also $\|g\|_{(p,r)} \leq \| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})} \|f\|_{(p,r)}$. Die Reduktion des Integrals über $[0,s]$ auf das Integral über $[0,1]$ ist möglich, da $\|\vec{f}(\cdot)\|_{\ell_q}$ periodisch mit Periode 1 ist.

Die gleiche Abschätzung gilt auch für die Supremumsnorm, d.h. bei $r = \infty$, denn

$$\begin{aligned}\|g\|_{(q,\infty)} := \sup_{x\in[0,1]} \|\vec{g}(x)\|_{\ell_q} &\leq \| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})} \sup_{y\in[0,s]} \|\vec{f}(y)\|_{\ell_q} \\ &= \| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})} \|f\|_{(q,\infty)} \,. \end{aligned} \tag{6.12}$$

Genau wie oben in Ungleichung (6.11) konnte hier das Supremum über das Intervall $[0,s]$ auf das Intervall $[0,1]$ eingeschränkt werden, da $\|\vec{f}(\cdot)\|_{\ell_q}$ periodisch mit Periode 1 ist.

Da $\|\cdot\|_{(q,r)}$ stetig auf $L^{q,r}(\mathbb{Z}\times\mathbb{T})$ und $\| \left(\downarrow\!\downarrow \mathrm{s}\right) \cdot\|_{(1,\tilde{q})}$ stetig auf $\ell_1(\mathbb{C})$ ist, und die Ungleichung 6.11 bzw. 6.12 für die jeweils dichten Teilmengen $C_c(\mathbb{R})$ und $\ell_{\text{fin}}(\mathbb{C})$ gezeigt wurde, gilt diese Ungleichung überall. Somit ist $\| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})}$ eine Schranke für den Operator $\mathcal{D}_s p(\mathcal{T})$. □

Satz 6.2.6 *Seien $q \in [1,\infty)$, $s \in \mathbb{N}_{>1}$ und $A \in \mathbb{N}_{>0}$ fixiert.*

Sei $p \in \ell_{1,A}(\mathbb{Z})$ eine Folge mit $\sum_{n\in\mathbb{Z}} p_n = 1$ und

$$\| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})} := \left\| \left(\|p_{(0)}\|_{\ell_1}, \|p_{(1)}\|_{\ell_1}, \ldots, \|p_{(s-1)}\|_{\ell_1} \right) \right\|_{\tilde{q}} \leq s^{A-1} \,,$$

wobei $\|\cdot\|_{\tilde{q}}$ die $\tilde{q}$–Norm auf $\mathbb{R}^s$ mit $\tilde{q} = 1 + \frac{1}{q-1}$ ist.

Dann gibt es genau eine stetige Funktion $\varphi \in L^q(\mathbb{R})$, die die Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ zur Skalierungsfolge $a = sH_s(\mathcal{T})^A p \in \ell_1(\mathbb{C})$ löst und als $\varphi = c(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A\eta$ mit $c \in \ell_{\text{fin}}(\mathbb{C})$ und $\eta \in L^{q,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^q(\mathbb{R}) \cap C_b(\mathbb{R})$ darstellbar ist.

Beweis: Dieser Satz ist eine Folgerung aus dem generischen Existenzsatz 6.1.5. Aus Lemma 6.2.4 folgt die Inklusion $C_c(\mathbb{R}) \subset L^{q,q}(\mathbb{Z}\times\mathbb{T}) \subset L^q(\mathbb{R})$, mit $q < \infty$ also $L^{q,q}(\mathbb{Z}\times\mathbb{T}) = L^q(\mathbb{R})$. Da nach Lemma 6.2.5 und mit der Voraussetzung $s^{1-A} \| \left(\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})} < 1$ der Operator $s^{1-A}\mathcal{D}_s p(\mathcal{T})$ kontraktiv auf $L^q(\mathbb{R})$ ist, gibt es eine Lösung der Verfeinerungsgleichung und genau eine Lösung des geforderten Typs.

Um die Stetigkeit dieser Lösung nachzuweisen, benutzen wir die gemischte (q,∞)–Norm. Lemma 6.2.5 besagt, dass $s^{1-A}\mathcal{D}_s p(\mathcal{T})$ kontraktiv auf $L^{q,\infty}(\mathbb{Z}\times\mathbb{T})$ mit Kontraktionskonstante

$$s^{1-A} \| \left(\downarrow \mathrm{s}\right) p\|_{(1,\tilde{q})} < 1$$

ist. Nach der Konstruktion im Beweis des Existenzsatzes 6.1.5 gibt es $c \in \ell_{\text{fin}}(\mathbb{C})$ und $\eta \in L^{q,\infty}(\mathbb{Z}\times\mathbb{T})$, so dass $\varphi = c(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A\eta$ eine Lösung der Verfeinerungsgleichung

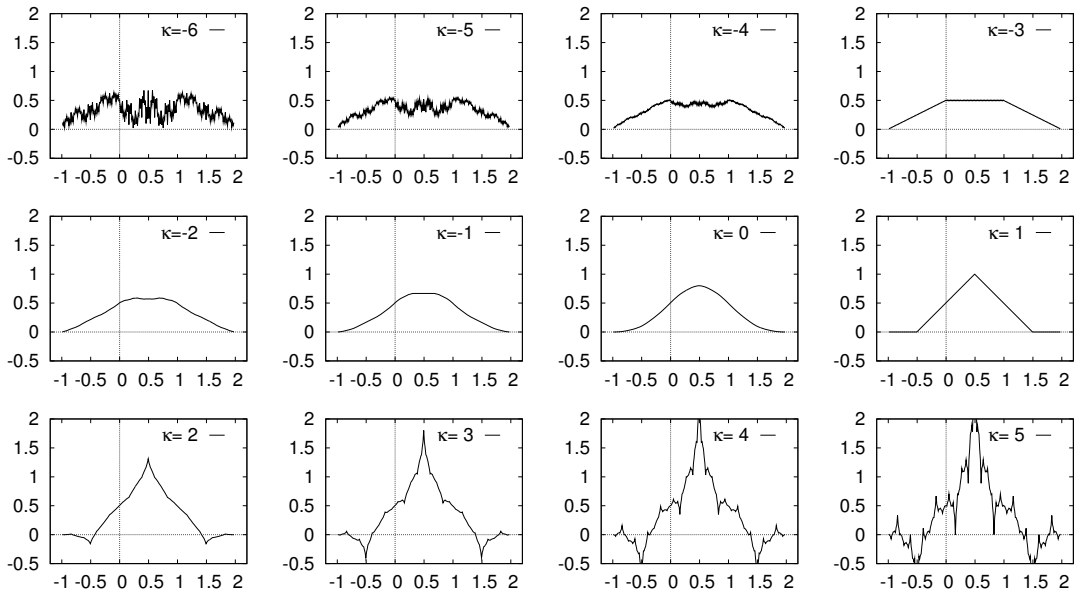


Abbildung 6.1: Symmetrische Skalierungsfunktionen, Dilatation 3 und Ordnung 2

$\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ ist. Nach Lemma 6.2.4 gilt $L^{q,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^q(\mathbb{R}) \cap C_b(\mathbb{R})$. Somit muss jede so konstruierte Lösung mit der oben gefundenen eindeutigen Lösung aus $L^q(\mathbb{R})$ übereinstimmen. □

Beispiel: Wir betrachten symmetrische Skalierungsfunktionen mit Skalenfaktor $s = 3$ und polynomialer Approximationsordnung $A = 2$. Das Laurent–Polynom zur Skalierungsfolge sei also $a(Z) = 3\,H_3(Z)^2\,p(Z)$ mit einem quadratischen Polynom

$$p(Z) = \frac{1}{4}((1+Z)^2 - \kappa(1-Z)^2) = \frac{1}{4}((1-\kappa) + 2(1+\kappa)Z + (1-\kappa)Z^2)\,.$$

Die Gleichung $p(Z)c(Z) - c(Z^3) = (1-Z)^2\tilde{c}(Z)$ hat bei $c(1) = 1$ eine Lösung $c(Z) = \frac{1}{2}(1+Z)$, $\tilde{c}(Z) = -\frac{3+\kappa}{8}(1+Z)$. Damit erhalten wir die affin lineare Fixpunktgleichung

$$\eta = -\frac{3+\kappa}{24}\mathcal{D}_3(1+\mathcal{T})\beta_1 + \frac{1}{3}\mathcal{D}_2 p(\mathcal{T})\eta\,,$$

die, wenn lösbar, auf die Lösung $\varphi = \frac{1}{2}(1+\mathcal{T})\beta_1 + (1-\mathcal{T})^2\eta$ der Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ führt.

Mit den bisherigen Überlegungen können wir u.a. folgende Aussagen zur Existenz und Beschaffenheit von Lösungen der Verfeinerungsgleichung machen:

i) Mit $\|p\|_1 = \frac{1}{2}(|1-\kappa| + |1+\kappa|) = \max(1, |\kappa|)$ erhalten wir Lösungen in $L^1(\mathbb{R})$ für $|\kappa| < s^{A-1+1} = 9$.

ii) Mittels der gleichen Norm, aber der kleineren Schranke $|\kappa| < s^{A-1+\frac{1}{2}} = 3\sqrt{3}$ erhalten wir Lösungen in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

iii) Wir erhalten stetige, beschränkte Lösungen in $L^1(\mathbb{R})$, wenn

$$\|\left(\downarrow\downarrow 3\right) p\|_{(1,\infty)} = \max(|p_0|, |p_2|, |p_1|) = \frac{1}{4}\max(|1-\kappa|, 2|1+\kappa|) < s^{A-1} = 3$$

ist, was für $-7 < \kappa < 5$ erfüllt ist. Wie wir noch sehen werden, erhalten wir für $-3 < \kappa < 1$ Lösungen, die darüber hinaus stetig differenzierbar sind.

iv) Wir erhalten stetige, beschränkte Lösungen in $L^2(\mathbb{R})$, wenn

$$\| \left(\downdownarrows 3\right) p\|_{(1,2)} = \sqrt{|p_0|^2 + |p_2|^2 + |p_1|^2} = \sqrt{\tfrac{1}{24}(8 + (1+3\kappa)^2)} < s^{A-1} = 3$$

ist, was für $-3 < \kappa < \frac{7}{3}$ gegeben ist.

6.2.3 Hölder–Stetigkeit der Lösungen

Definition 6.2.7 *Eine Funktion $f : \mathbb{R} \to \mathbb{C}$ heißt* Hölder–stetig *zum Index $\alpha \in (0,1]$, wenn es eine Konstante $C > 0$ gibt, so dass*

$$\left|f(y) - f(x)\right| \leq C\,|y-x|^{\alpha}$$

für beliebige $x, y \in \mathbb{R}$ gilt.

Somit ist jede Hölder–stetige Funktion auch gleichmäßig stetig.

Satz 6.2.8 *Sei eine Skalierungsfolge $a := sH_s(\mathcal{T})^A p$ mit $p \in \ell_{1,A}(\mathbb{Z})$ gegeben, und sei*

$$\| \left(\downdownarrows \mathrm{s}\right) p\|_{(1,\infty)} = \max_{0\leq j<s} \sum_{k\in\mathbb{Z}} |p_{sk+j}| = s^{A-1-\alpha^*}$$

mit einem $\alpha^* \in (0,1]$. Dann hat die Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ eine eindeutig bestimmte Lösung der Form $\varphi = c(\mathcal{T})\beta_m + (1-\mathcal{T})^A\eta$ mit $m \geq A-1$, $c \in \ell_{\mathrm{fin}}(\mathbb{C})$ und $\eta \in L^1(\mathbb{R})$. Ist $\alpha^* < 1$, so ist φ Hölder–stetig mit Index α^*, ist $\alpha^* = 1$, so ist φ Hölder–stetig für jeden Index $\alpha \in (0,1)$.

Beweis: Nach Satz 6.2.6 gibt es unter diesen Voraussetzungen eine stetige, beschränkte Lösung $\varphi \in L^1(\mathbb{R}) \cap C_b(\mathbb{R})$.

Wir können, durch eine leichte Erweiterung der Konstruktion im Beweis des Existenzsatzes 6.1.5, die Darstellung der eindeutigen bestimmten Skalierungsfunktion auf die Form $\varphi = c(\mathcal{T})\beta_m + (1-\mathcal{T})^A\eta$ erweitern, mit der Ordnung $m \geq \max(2, A-1)$ des B–Splines β_m. Dabei sind $c \in \ell_{\mathrm{fin}}(\mathbb{C})$ eine endliche Folge und $\eta \in L^1(\mathbb{C}) \cap C_b(\mathbb{R})$ die Lösung der affin linearen Fixpunktgleichung

$$\eta = s^{1-A}\mathcal{D}_s\Big(\tilde{c}(\mathcal{T})\beta_m + p(\mathcal{T})\eta\Big)\,. \tag{6.13}$$

Das Paar $(c, \tilde{c}) \in \ell_{\mathrm{fin}}(\mathbb{C}) \times \ell_1(\mathbb{C})$ muss dazu eine Lösung der Gleichung

$$p(\mathcal{T})c - \tilde{p}(\mathcal{T})\left(\uparrow \mathrm{s}\right)c = (1-\mathcal{T})\tilde{c} \tag{6.14}$$

mit $\tilde{p}(Z) = H_s(Z)^{m+1-A}$ sein. Da die Folge $\tilde{p}$ endlich ist und $\tilde{p}(1) = 1$ gilt, gibt es nach Satz 6.1.3 ein solches Paar $(c, \tilde{c})$. Für die $(1,\infty)$–Norm von η erhalten wir aus der Fixpunktgleichung die Abschätzung

$$\|\eta\|_{(1,\infty)} \leq \frac{1}{1-s^{-\alpha^*}}\; s^{1-A}\| \left(\downdownarrows \mathrm{s}\right)\tilde{c}\|_{(1,\infty)}\,\|\beta_m\|_{(1,\infty)}\,. \tag{6.15}$$

Um die Hölder–Stetigkeit von φ nachzuweisen, ist es ausreichend, η zu betrachten. Denn die Funktion $c(\mathcal{T})\beta_m$ ist wegen $c \in \ell_{\text{fin}}(\mathbb{C})$ und $m \geq 2$ stetig differenzierbar und hat kompakten Träger. Daher ist diese Funktion auch zu jedem Index $\alpha \in (0,1]$ Hölder–stetig. Um nun die Hölder–Stetigkeit von η zu beweisen, müssen wir Differenzen von Funktionswerten abschätzen. Im Rahmen dieser Theorie geschieht dies, indem die Differenz zweier Wertefolgen $\vec{\eta}(y) - \vec{\eta}(x)$ abgeschätzt wird. Dabei war $\vec{f}(x) := \{f(x+n)\}_{n\in\mathbb{Z}}$ für beliebige stetige Funktionen $f \in C(\mathbb{R})$ definiert.

Die Fixpunktgleichung von η kann als Beziehung der Wertefolgen geschrieben werden. Für die Differenz der Wertefolgen zu zwei verschiedenen Argumenten gilt dann

$$\vec{\eta}(y) - \vec{\eta}(x) = \left(\downarrow \mathrm{s}\right) \left(\tilde{c}(\mathcal{T}) \left(\vec{\beta}_m(sy) - \vec{\beta}_m(sx)\right) + p(\mathcal{T}) \left(\vec{\eta}(sy) - \vec{\eta}(sx)\right)\right) . \tag{6.16}$$

Um die Größe dieser Differenz abzuschätzen, betrachten wir die zwei Folgen $\{\varepsilon_n\}_{n\in\mathbb{Z}}, \{b_n\}_{n\in\mathbb{Z}} \subset \mathbb{R}$ mit Gliedern

$$\begin{aligned}
\varepsilon_n :=& s^{\alpha^* n} \sup\{\|\vec{\eta}(x) - \vec{\eta}(y)\|_{\ell_1} : \ x, y \in \mathbb{R},\ |x-y| \in [s^{-n}, s^{1-n}]\} \\
b_n :=& s^{\alpha^* n} \sup\{\|\vec{\beta}_m(x) - \vec{\beta}_m(y)\|_{\ell_1} : \ x, y \in \mathbb{R},\ |x-y| \in [s^{-n}, s^{1-n}]\} .
\end{aligned}$$

Wir werden nachfolgend zeigen, dass $\{\varepsilon_n\}_{n\in\mathbb{Z}}$ für $\alpha^* < 1$ durch ein E_{α^*} beschränkt ist und linear wachsend für $\alpha^* = 1$. Dann gibt es bei $\alpha^* = 1$ für jedes $\alpha \in (0,1)$ eine Konstante $E_\alpha \in \mathbb{R}$, so dass $\{s^{(\alpha-\alpha^*)n}\varepsilon_n\}_{n\in\mathbb{Z}}$ durch dieses E_α beschränkt ist. Damit kann die Differenz der Funktionswerte zu jedem Paar $x, y \in \mathbb{R}$ mit Abstand $|x-y| < 1$ abgeschätzt werden als

$$|\eta(x) - \eta(y)| \leq \|\vec{\eta}(x) - \vec{\eta}(y)\|_{\ell_1} \leq s^{-\alpha^* n}\varepsilon_n \leq E_\alpha s^{-\alpha n} \leq E_\alpha |x-y|^\alpha .$$

Dabei ist $n \in \mathbb{Z}$ dadurch bestimmt, dass $s^{-1}|x-y| < s^{-n} \leq |x-y|$ gelten soll. Somit ist η unter obiger Annahme Hölder–stetig mit Index α.

Seien nun ein $n \in \mathbb{Z}$ und zwei Punkte $x, y \in \mathbb{R}$ mit $s^{-n-1} \leq |x-y| < s^{-n}$ fixiert. Analog zu (6.9) können wir den Abstand der Folgen $\vec{\eta}(x), \vec{\eta}(y) \in \ell_1(\mathbb{Z})$ abschätzen als

$$\begin{aligned}
s^{(n+1)\alpha^*}\|\vec{\eta}(x) - \vec{\eta}(y)\|_{\ell_1} \leq \quad & s^{1-A+\alpha^*} \| \left(\downarrow\downarrow \mathrm{s}\right) \tilde{c}\|_{(1,\infty)} \, s^{n\alpha^*} \left\|\vec{\beta}_m(sx) - \vec{\beta}_m(sy)\right\|_{\ell_1} \\
& + s^{1-A+\alpha^*} \| \left(\downarrow\downarrow \mathrm{s}\right) p\|_{(1,\infty)} \, s^{n\alpha^*} \, \|\vec{\eta}(sx) - \vec{\eta}(sy)\|_{\ell_1} , \\
\text{d.h. } \varepsilon_{n+1} \leq & s^{1-A+a^*} \| \left(\downarrow\downarrow \mathrm{s}\right) \tilde{c}\|_{(1,\infty)} b_n + \varepsilon_n .
\end{aligned}$$

Durch Induktion über n erhält man daraus

$$\varepsilon_n \leq s^{1-A+a^*} \| \left(\downarrow\downarrow \mathrm{s}\right) \tilde{c}\|_{(1,\infty)} \sum_{k=1}^{n-1} b_k + \varepsilon_1$$

mit der Abschätzung des Startwerts

$$\varepsilon_1 \leq 2s^{a^*}\|\eta\|_{(1,\infty)} \leq 2\| \left(\downarrow\downarrow \mathrm{s}\right) \tilde{c}\|_{(1,\infty)} \frac{s^{1-A+\alpha^*}}{1 - s^{-\alpha^*}} .$$

Wir brauchen also noch Abschätzungen für die Glieder der Folge $\{b_n\}_{n\in\mathbb{Z}}$. Der B–Spline β_m hat die Ableitung $\beta_m' = (1-\mathcal{T})\beta_{m-1}$ und es gelten sowohl $\beta_{m-1}(x) > 0$ als auch $\sum_{n\in\mathbb{Z}} \beta_m(x+n) = 1$. Wegen $m-1 > 0$ ist β_{m-1} stetig und hat kompakten Träger, damit erhalten wir für $x < y$

$$\begin{aligned}\left\|\vec{\beta}_m(x) - \vec{\beta}_m(y)\right\|_{\ell_1} &= \sum_{n\in\mathbb{Z}} \left| \int_{x+n}^{y+n} (1-\mathcal{T})\beta_{m-1}(\tau)\,\mathrm{d}\tau \right| \\ &\le \int_x^y \sum_{n\in\mathbb{Z}} \left(|\beta_{m-1}(\tau+n)| + |\beta_{m-1}(\tau+n-1)|\right) \mathrm{d}\tau = 2|x-y| \,.\end{aligned}$$

Daraus folgt $b_n \le 2s^{n\alpha^*} s^{1-n} = 2s^{1-n(1-\alpha^*)}$ für alle $n \in \mathbb{N}$. Für $\alpha^* < 1$ erhalten wir somit $b_n \le 2s^{\alpha^* n}\, s^{-n}$ und

$$\sum_{k=1}^{n-1} b_k \le 2s \sum_{k=1}^{n-1} s^{(\alpha^*-1)k} = 2s \frac{s^{\alpha^*-1} - s^{(\alpha^*-1)n}}{1 - s^{\alpha^*-1}} \le 2\frac{s^{\alpha^*}}{1-s^{\alpha^*-1}} \,,$$

und daher

$$\varepsilon_n \le E_{\alpha^*} := 2s^{1-A+\alpha^*} \| \left(\downarrow\!\!\downarrow \mathrm{s}\right) \tilde{c}\|_{(1,\infty)} \left(\frac{s^{\alpha^*}}{1-s^{\alpha^*-1}} + \frac{1}{1-s^{-\alpha^*}} \right) .$$

Für $\alpha^* = 1$ hingegen ergibt sich

$$\sum_{k=1}^{n-1} b_k \le s \sum_{k=1}^{n-1} s^0 = (n-1)s$$

und damit

$$\varepsilon_n \le 2s^{1-A} \| \left(\downarrow\!\!\downarrow \mathrm{s}\right) \tilde{c}\|_{(1,\infty)} \left(s^2(n-1) + \frac{s}{s-1} \right) .$$

□

6.2.4 Differenzierbarkeit der Lösungen

Ist $s^{-\alpha}$ eine Schranke des Operators $s^{1-A}\mathcal{D}_s p(\mathcal{T})$ in einem der Räume $L^{q,r}(\mathbb{Z}\times\mathbb{T})$, so haben wir eben gesehen, dass die Differenz von α zu Null sich in eine Güte der Stetigkeit der Lösung der Skalierungsgleichung $\varphi = s\mathcal{D}_s H_s(\mathcal{T})^A p(\mathcal{T})\varphi$ übersetzen läßt, sofern $\alpha \in (0,1]$ gilt. Ist $\alpha > 1$, so kann aus dem ganzzahligen Anteil von α sogar auf die stetige Differenzierbarkeit von φ gefolgert werden.

Lemma 6.2.9 *Sind $\varphi_1, \varphi_2 \in L^1(\mathbb{R})$ Lösungen von Verfeinerungsgleichungen $\varphi_k = \mathcal{D}_s(a_k(\mathcal{T})\varphi_k)$, $k = 1,2$ mit Skalierungsfolgen $a_1, a_2 \in \ell_1(\mathbb{Z})$, so ist auch deren Faltungsprodukt $\varphi := \varphi_1 * \varphi_2 \in L^1(\mathbb{R})$ Lösung einer Verfeinerungsgleichung $\varphi = \mathcal{D}_s(a(\mathcal{T})\varphi)$, wobei sich die Skalierungsfolge a als Faltungsprodukt $a = \frac{1}{s} a_1 * a_2 \in \ell_1(\mathbb{Z})$ ergibt.*

Beweis: Nach Voraussetzung können im folgenden Reihenbildung und Integration vertauscht werden, denn die Reihen sind absolut summierbar und die Integranden absolut integrierbar.

$$\begin{aligned}(\varphi_1 * \varphi_2)(x) &= \int_{\mathbb{R}} \varphi_1(t)\varphi_2(x-t)\,\mathrm{d}t = \sum_{k,l\in\mathbb{Z}} \int_{\mathbb{R}} a_{1,k}\varphi_1(st-k)a_{2,l}\varphi_2(sx-st-l)\,\mathrm{d}t \\ &= \sum_{n\in\mathbb{Z}} \left(\sum_{k\in\mathbb{Z}} a_{1,k}a_{2,n-k}\right) \int_{\mathbb{R}} \varphi_1(st-k)\varphi_2(sx-st-n+k)\,\mathrm{d}t \\ &= \sum_{n\in\mathbb{Z}} (a_1 * a_2)_n \int_{\mathbb{R}} \varphi_1(y)\varphi_2(sx-n-y)\,\frac{1}{s}\,\mathrm{d}y = \sum_{n\in\mathbb{Z}} a_n\varphi(sx-n)\,.\end{aligned}$$

□

Satz 6.2.10 *Sei eine Skalierungsfolge $a(Z) := sH_s(Z)^A p(Z)$ mit $p \in \ell_{1,A}(\mathbb{Z})$ gegeben, und sei*

$$\|\,(\downarrow\!s)\,p\|_{1,\infty} = \max_{0\le j<s} \sum_{k\in\mathbb{Z}} |p_{sk+j}| = s^{A-1-r-\alpha^*}$$

mit einem $r \in \mathbb{N}$ und $\alpha^* \in (0,1]$. Dann hat die Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ eine eindeutig bestimmte, r–fach stetig differenzierbare Lösung, deren r–te Ableitung für jeden Hölder–Index $\alpha \in (0,\alpha^*] \cap (0,1)$ Hölder–stetig ist.

Beweis: Nach Satz 6.2.8 hat die Verfeinerungsgleichung $\varphi_r = s\mathcal{D}_s H_s(Z)^{A-r} p(Z)\varphi_r$ eine Hölder–stetige Lösung mit Hölder–Index α^* falls $\alpha^* < 1$ und $\alpha \in (0,1)$ falls $\alpha^* = 1$.

Nach Lemma 6.2.9 ist das Faltungsprodukt $\varphi := \beta_{r-1} * \varphi_r$ eine Lösung von $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$. Das Faltungsprodukt einer k–fach stetig differenzierbaren Funktion f mit $\beta_0 = \chi_{[0,1)}$ ergibt eine $k+1$–fach stetig differenzierbare Funktion, denn es gilt $(f * \beta_0)(x) = F(x) - F(x-1)$ für eine Stammfunktion F von f. Daher ist φ r–fach stetig differenzierbar, denn diese Funktion entsteht durch r–fache Faltung mit β_0. □

6.3 Existenz biorthogonaler Paare von Skalierungsfunktionen

Wir wissen nun, unter welchen Bedingungen es zu einer Skalierungsfolge eine Lösung der Verfeinerungsgleichung gibt, die stetig ist und ein verschiebungsinvariantes Bessel–System erzeugt. Damit diese Lösung eine Multiskalenanalyse erzeugt, muss dieses verschiebungsinvariante System nach Satz 5.3.4 auch ein Riesz–System sein. Paradoxerweise ist dies am einfachsten nachzuweisen, wenn dieser Nachweis für ein Paar von Skalierungsfolgen mit Skalierungsfunktionen gleichzeitig geführt wird.

Denn ist ein Paar zulässiger Skalierungsfunktionen (s. Definition 5.3.3) biorthogonal (s. Definition 5.3.11), so sind diese nach Satz 5.3.13 zulässig und erzeugen jeweils eine Multiskalenanalyse. Die Biorthogonalität der Skalierungsfunktionen kann, wie nachfolgend gezeigt wird, auf die Biorthogonalität der Skalierungsfolgen zurückgeführt werden. Zu einer gegebenen endlichen Skalierungsfolge eine weitere endliche Folge zu finden, die diese zu einem biorthogonalen Paar ergänzt, entspricht der Lösung eines linearen Gleichungssystems. Dass diese

zweite Folge die Lösbarkeitsbedingung der Verfeinerungsgleichung erfüllt, ist weniger leicht zu erreichen.

6.3.1 Existenz zulässiger Skalierungsfunktionen

Die Betrachtung der schwächsten gemischten Norm in der Lösbarkeitstheorie der Verfeinerungsgleichung ist nicht nur ausreichend, um auf eine stetige Lösung zu schließen, sondern genügt auch, um eine zulässige Skalierungsfunktion zu erhalten.

Satz 6.3.1 *Seien $s \in \mathbb{N}_{>1}$ und $A \in \mathbb{N}_{>0}$ fixiert. Sei $p \in \ell_{1,A}(\mathbb{Z})$ eine Folge mit $\sum_{n\in\mathbb{Z}} p_n = 1$ und gelte*

$$\| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\infty)} < s^{A-1} \, .$$

Für die durch $a(\mathcal{T}) = sH_s(\mathcal{T})^A p(\mathcal{T})$ definierte Skalierungsfolge $a \in \ell_1(\mathbb{C})$ gibt es eine eindeutige Lösung $\varphi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \cap C_b(\mathbb{R})$ der Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$, welche eine zulässige Skalierungsfunktion ist und eine Approximationsbedingung der Ordnung $A-1$ erfüllt.

Beweis: Wegen $s^{1-A}\| \left(\downarrow\!\downarrow \mathrm{s}\right) p\|_{(1,\infty)} < s^1$ gibt es nach Satz 6.2.6 eine eindeutige Lösung $\varphi \in L^{1,\infty}(\mathbb{Z}\times\mathbb{T})$ der Form $\varphi = c(\mathcal{T})\beta_{A-1} + (1-\mathcal{T})^A\eta$ mit $c \in \ell_{\mathrm{fin}}(\mathbb{C})$ und $\eta \in L^{1,\infty}(\mathbb{Z}\times\mathbb{T})$. Nach Lemma 6.2.4 gelten damit folgende Einbettungen

$$\eta, \varphi \in L^{1,\infty}(\mathbb{Z}\times\mathbb{T}) \subset L^{1,\infty}(\mathbb{Z}\times\mathbb{T}) \cap L^{2,\infty}(\mathbb{Z}\times\mathbb{T}) \subset C_b(\mathbb{R}) \cap L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \, .$$

Um zu zeigen, dass φ eine zulässige Skalierungsfunktion ist, genügt es zu zeigen, dass η beschränkte Prä–Gramsche Fasern hat. Dies erfolgt mit einer Version der *Poissonschen Summenformel*.

Sei dazu $\alpha \in \mathbb{R}$ fixiert. Es gilt $\eta \in L^{1,\infty}(\mathbb{Z}\times\mathbb{T}) \subset C_b(\mathbb{R})$, d.h. η ist stetig und die Periodisierung $g_\alpha := \sum_{k\in\mathbb{Z}} \mathcal{T}^k(e_\alpha\eta)$ konvergiert überall absolut. Denn mit der gemischten $(1,\infty)$–Norm gilt für jedes $x \in \mathbb{R}$

$$|g_\alpha(x)| \le \sum_{k\in\mathbb{Z}} |\eta(x+k)| \le \|\eta\|_{(1,\infty)} < \infty \, .$$

Somit ist g_α messbar, beschränkt und periodisch mit Periode 1, besitzt also eine Entwicklung in eine Fourier–Reihe über dem Intervall $I := [0,1]$,

$$g_\alpha = \sum_{n\in\mathbb{Z}} \langle\, g_\alpha,\, e_n \,\rangle_{L^2(I)}\, e_n \, .$$

Die Fourier–Koeffizienten ergeben sich zu

$$\begin{aligned} \langle\, g_\alpha,\, e_n \,\rangle &= \int_0^1 \sum_{k\in\mathbb{Z}} e_\alpha(x-k)\eta(x-k)e_{-n}(x)\,\mathrm{d}x = \sum_{k\in\mathbb{Z}} \int_k^{k+1} e_{\alpha-n}(x)\eta(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}} \eta(x) e^{-i\,2\pi(n-\alpha)\,x}\,\mathrm{d}x = \hat{\eta}(n-\alpha) \end{aligned}$$

als Werte der Fourier–Transformierten $\hat{\eta} := \mathcal{F}(\eta) \in L^2(\mathbb{R}) \cap C_b(\mathbb{R})$. Somit gilt für jedes $x \in \mathbb{R}$ die Poissonsche Summenformel

$$e_\alpha(x) \sum_{k\in\mathbb{Z}} \eta(x+k)e_k(\alpha) = g_\alpha(x) = \sum_{n\in\mathbb{Z}} \hat{\eta}(n-\alpha)e_n(x) \, .$$

Die ℓ_2–Norm der Prä–Gramschen Faser $J_\eta(-\alpha)$ ist die Reihe der Betragsquadrate der Fourier–Koeffizienten der Fourier–Entwicklung von g_α, es gilt

$$\|J_\eta(-\alpha)\|_{\ell_2}^2 = \|g_\alpha\|_{L^2(I)}^2 = \int_0^1 \left| \sum_{k\in\mathbb{Z}} \eta(x+k) e_k(\alpha) \right|^2 \mathrm{d}x \leq \|\eta\|_{(1,\infty)}^2 < \infty \,.$$

Für die Prä–Gramsche Faser von φ erhalten wir daraus

$$\left\| J_\varphi(\omega) - \hat{c}(\omega) J_{\beta_{A-1}}(\omega) \right\|_{\ell_2} = \left\| (1-\eta_1(\omega))^A J_\eta(\omega) \right\|_{\ell_2} \leq |\sin(\pi\omega)|^A \|\eta\|_{(1,\infty)} \,,$$

also in Landau–Notation $J_\varphi - \hat{c} J_{\beta_{A-1}} = O_{\ell_2}(\omega^A)$, folglich auch

$$J_\varphi - \hat{c} J_{\beta_{A-1}} = o_{\ell_2}(\omega^{A-1}) \,.$$

Nach Satz 5.3.8 ist dies eine äquivalente Formulierung der Approximationsbedingung der Ordnung $A-1$. □

Lemma 6.3.2 *Seien $s \in \mathbb{N}_{>1}$, $A \in \mathbb{N}_{>0}$ und $p \in \ell_{1,A}(\mathbb{C})$ eine Folge mit $\sum_{n\in\mathbb{Z}} p_n = 1$ und $\| \left(\downarrow\!\!\downarrow \mathrm{s}\right) p\|_{(1,\infty)} < s^{A-1}$.*

Gibt es eine endliche Folge $v \in \ell_{\mathrm{fin}}(\mathbb{C})$ mit $\sum_{n\in\mathbb{Z}} v_n = 1$, welche die Gleichung $v = \left(\downarrow \mathrm{s}\right) a(\mathcal{T}) v$ mit der Skalierungsfolge $a = s H_s(\mathcal{T})^A p$ erfüllt, so stimmt v mit der Wertefolge $\vec{\varphi}(0)$ der zu a gehörigen stetigen Skalierungsfunktion φ überein.

Beweis: Sei ein $m \in \mathbb{N}$ mit $m \geq \min(1, A-1)$ beliebig gewählt. Die Wertefolge $\vec{\beta}_m(0) = \{\beta_m(n)\}_{n\in\mathbb{Z}}$ des B–Splines der Ordnung m ist endlich und hat die Summe 1. Daher ist das Polynom $\sum_{n=0}^{m+1} \beta_m(n) Z^n$ modulo $(1-Z)^A$ nach dem erweiterten Euklidischen Algorithmus invertierbar. Es existiert also eine endliche Folge c, mit welcher

$$v \equiv c(\mathcal{T}) \vec{\beta}_m(0) \bmod (1-\mathcal{T})^A \ell_{\mathrm{fin}}(\mathbb{C})$$

gilt. Nach Satz 3.5.3 folgt aus dem Bestehen der Gleichung $\left(\downarrow \mathrm{s}\right)(H_s(\mathcal{T})^A w) = v$ für die endliche Folge $w = s p(\mathcal{T}) v \in \ell_{\mathrm{fin}}(\mathbb{C})$ die Beziehung

$$\frac{1}{s} H_s(\mathcal{T})^A w \equiv \left(\uparrow \mathrm{s}\right) v \bmod (1-\mathcal{T})^A \ell_{\mathrm{fin}}(\mathbb{C}) \,.$$

Für die Wertefolge des B–Splines gilt die Verfeinerungsgleichung

$$\vec{\beta}_m(0) = s \left(\downarrow \mathrm{s}\right) H_s(\mathcal{T})^{m+1} \vec{\beta}_m(0) \,,$$

daraus folgt analog zu oben

$$H_s(\mathcal{T})^{m+1} \vec{\beta}_m(0) \equiv \left(\uparrow \mathrm{s}\right) \vec{\beta}_m(0) \bmod (1-\mathcal{T})^A \ell_{\mathrm{fin}}(\mathbb{C}) \,.$$

Aus der Kombination der drei Gleichungen folgt

$$\begin{aligned} H_s(\mathcal{T})^{m+1-A} c(\mathcal{T}^s) \left(\uparrow \mathrm{s}\right) \vec{\beta}_m(0) \equiv & H_s(\mathcal{T})^{m+1-A} \left(\uparrow \mathrm{s}\right) v \equiv H_s(\mathcal{T})^{m+1} p(\mathcal{T}) c(\mathcal{T}) \vec{\beta}_m(0) \\ \equiv & p(\mathcal{T}) c(\mathcal{T}) \left(\uparrow \mathrm{s}\right) \vec{\beta}_m(0) \,. \end{aligned}$$

Da die Wertefolge $\vec{\beta}_m(0)$ modulo $(1-\mathcal{T})^A \ell_{\text{fin}}(\mathbb{C})$ invertierbar ist, muss $H_s(\mathcal{T})^{m+1-A}\left(\uparrow \mathrm{s}\right) c \equiv p(\mathcal{T})c$ gelten. Es gibt also eine endliche Folge $\tilde{c}$, mit welcher die Identität

$$p(\mathcal{T})c - H_s(\mathcal{T})^{m+1-A}\left(\uparrow \mathrm{s}\right) c = (1-\mathcal{T})^A \tilde{c}$$

erfüllt ist. Ist weiter w diejenige endliche Folge, mit welcher $v = c(\mathcal{T})\vec{\beta}_m(0) + (1-\mathcal{T})^A w$ gilt, so erfüllt w analog zur Konstruktion in Existenzsatz 6.1.5 die Fixpunktgleichung

$$w = s^{1-A}\left(\downarrow \mathrm{s}\right)\left(\tilde{c}(\mathcal{T})\vec{\beta}_m(0) + p(\mathcal{T})w\right) .$$

Dies ist jedoch auch die Fixpunktgleichung für die Wertefolge $\vec{\eta}(0)$ der Lösung $\eta \in \mathcal{L}^1(\mathbb{R}) \cap L^2(\mathbb{R}) \cap C(\mathbb{R})$ der nach Voraussetzung kontraktiven Fixpunktgleichung

$$\eta = s^{1-A}\mathcal{D}_s\left(\tilde{c}(\mathcal{T})\beta_m + p(\mathcal{T})\eta\right) .$$

Daher gilt $w = \vec{\eta}(0)$ und, da $\varphi := c(\mathcal{T})\beta_m + (1-\mathcal{T})^A\eta$ nach Satz 6.2.6 die eindeutige Lösung der Verfeinerungsgleichung $\varphi = \mathcal{D}_s a(\mathcal{T})\varphi$ ist, gilt auch $v = \vec{\varphi}(0)$. □

Satz 6.3.3 *Seien $s \in \mathbb{N}_{>1}$, $A, \tilde{A} \in \mathbb{N}_{>0}$ und $p \in \ell_{1,A}(\mathbb{C})$ sowie $\tilde{p} \in \ell_{1,\tilde{A}}(\mathbb{C})$ so gegeben, dass*

$$\|\left(\downdownarrows \mathrm{s}\right) p\|_{(1,\infty)} < s^{A-1} \text{ und } \|\left(\downdownarrows \mathrm{s}\right) \tilde{p}\|_{(1,\infty)} < s^{\tilde{A}-1}$$

und für die Differenzoperatoren $a(\mathcal{T}) := sH_s(\mathcal{T})^A p(\mathcal{T})$ und $\tilde{a}(\mathcal{T}) := sH_s(\mathcal{T})^{\tilde{A}}\tilde{p}(\mathcal{T})$ die Identität

$$\left(\downarrow \mathrm{s}\right) \circ \tilde{a}(\mathcal{T})^* \circ a(\mathcal{T}) \circ \left(\uparrow \mathrm{s}\right) = s\, id_{\ell_2(\mathbb{C})}$$

erfüllt ist.

Dann existieren zulässige stetige Skalierungsfunktionen φ und $\tilde{p}$hi zu den Skalierungsfolgen $a = sH_s(\mathcal{T})^A p$ und $\tilde{a} = sH_s(\mathcal{T})^{\tilde{A}}\tilde{p}$, mit Approximationsordnungen $A-1$ bzw.$\tilde{A}-1$, welche ein biorthogonales Paar bilden. Insbesondere erzeugen beide Funktionen eine Multiskalenanalyse.

Beweis: Dass die Verfeinerungsgleichungen zu den Folgen a und $\tilde{a}$ eindeutige stetige Lösungen der angegebenen Approximationsordnung haben, war Inhalt des vorhergehenden Satzes 6.3.1. Es verbleibt zu zeigen, dass diese beiden Funktionen in der angegebenen Art komplementär zueinander sind.

Dazu betrachten wir das Faltungsprodukt $\Phi := \tilde{\varphi}_- * \varphi$, wobei $\tilde{\varphi}_-(x) := \overline{\tilde{\varphi}(-x)}$ die konjugierte und zeitinvertierte Funktion zu $\tilde{\varphi}$ ist. Mit dieser gilt, da das Faltungsprodukt stetig ist,

$$\Phi(n) := (\tilde{\varphi}_- * \varphi)(n) = \int_{\mathbb{R}} \varphi(x)\overline{\tilde{\varphi}(x-n)}\,\mathrm{d}x = \langle\, \varphi,\, \mathcal{T}^n\tilde{\varphi}\,\rangle_{L^2} .$$

Nach Lemma 6.2.9 ist Ψ eine Lösung der Verfeinerungsgleichung $\Phi = s^{-1}\mathcal{D}_s\tilde{a}(\mathcal{T})^* a(\mathcal{T})\Phi$ mit Approximationsordnung $A+\tilde{A}-1$. Sei $P := \tilde{p}(\mathcal{T})^* p$ das Faltungsprodukt der Faktoren $p, \tilde{p}$ der Skalierungsfolgen. Die Polyphasennorm des Faltungsprodukts P besitzt die Abschätzung

$$\begin{aligned}
\|\left(\downdownarrows \mathrm{s}\right) P\|_{1,\infty} &= \max_{k=0,\ldots,s-1} \|\left(\downarrow \mathrm{s}\right)(\mathcal{T}^{-k}\tilde{a}(\mathcal{T})^* a)\|_{\ell_1} \\
&\leq \max_{k=0,\ldots,s-1} \|\left(\downdownarrows \mathrm{s}\right)\tilde{a}(\mathcal{T})^*\|_{1,\infty}\, \|\mathcal{T}^{-k}a\|_{\ell_1} \\
&< s^{\tilde{A}-1}\left(s\, s^{A-1}\right) = s^{A+\tilde{A}-1}
\end{aligned}$$

Daher ist Φ auch die einzige Lösung dieser Verfeinerungsgleichung. Insbesondere ist die Folge $\vec{\Phi}(0)$ die eindeutige Lösung der Gleichung

$$\vec{\Phi}(0) = \frac{1}{s}\,(\downarrow s)\left(a(\mathcal{T})^* a(\mathcal{T})\vec{\Phi}(0)\right).$$

Da nach Voraussetzung auch die Folge δ^0 diese Gleichung löst, muss nach Lemma 6.3.2 $\vec{\Phi}(0) = \delta^0$ gelten. Daraus folgt aber auch, dass die Verknüpfung $\mathcal{E}^*_{\tilde{\varphi}} \circ \mathcal{E}_{\varphi} : \ell_2(\mathbb{C}) \to \ell_2(\mathbb{C})$ die identische Abbildung ist, was nach Lemma 5.3.12 bedeutet, dass $(\varphi, \tilde{\varphi})$ ein biorthogonales Paar ist. □

6.4 Weitere Beispiele symmetrisch–orthogonaler Skalierungsfunktionen

Für jedes Tripel $(S, A, V) \in \mathbb{N}_{\geq 2} \times \mathbb{N}_{\geq 1} \times \mathbb{N}_0$ aus Skalenfaktor, Approximationsordnung und Anzahl freier Variablen kann man nach der Menge der Parametertupel in $\mathbb{R}^V$ suchen, welche eine stetige symmetrische orthogonale Skalierungsfunktion erzeugen.

6.4.1 Algorithmus zum Aufstellen des Gleichungssystems

Die Bedingungen, die die Parameter erfüllen müssen, ergeben sich wie folgt:

- Die Gleichungen, die zu konstruieren sind, sind Elemente des Polynomrings $\mathcal{R} := \mathbb{Q}[Y_1, \ldots, Y_V]$.
- Im Ring $\mathbb{Q}[[X]]$ formaler Potenzreihen werden die Polynome

$$\begin{aligned}
h_S(X) :=& 1 + \sum_{k=1}^{S-1} \prod m = 1^k \left(\frac{S^2 - m^2}{(m+1)(2m+1)}\right), (-X)^k \\
Q_{S,A}(X) :=& h_S(X)^{-A} \bmod X^A, \\
C_{S,A}(X) :=& \begin{cases} 1 - \frac{1}{2}X & \text{wenn } S \text{ und } A \text{ gerade} \\ 1 & \text{sonst,} \end{cases} \\
q_{S,A}(X) :=& \left(C_{S,A}(X)^{-1}\, Q_{S,A}(X)\right)^{\frac{1}{2}} \bmod X^A
\end{aligned}$$

  bestimmt. Da die Potenzreihen alle einen konstanten Koeffizienten 1 haben, sind sowohl die negative Potenz als auch die Quadratwurzel definiert und können rein arithmetisch bestimmt werden. Es sei weiter vereinbart, dass $Q_{S,A}$ und $q_{S,A}$ die kleinsten Repräsentanten modulo X^A seien, d.h. sie sind die nach dem Glied zum Grad $(A-1)$ abgebrochenen Potenzreihen. Somit können beide auch als Polynome interpretiert werden.

- Nach diesen Vorbereitungen wird der Ansatz

$$q(X) := q_{S,A}(X) + X^A(Y_1 + Y_2 X + \cdots + Y_V X^{V-1}) \in \mathcal{R}[X]$$

  gebildet. Nach Konstruktion ist $C_{S,A}(X)q(X)^2 - Q_{S,A}(X)$ ein Vielfaches von X^A, es gibt also ein Polynom $r \in \mathcal{R}[X]$, so dass $X^A r(X)$ gerade diese Differenz ist. Damit hat r den

Grad $\deg r = A + 2(V-1)$. Sind S und A beide gerade, so erhöht sich dieser Grad um Eins.

- Nun wird zum Ring $\mathcal{R}\langle Z\rangle$ der Laurent–Polynome mit Koeffizienten in $\mathcal{R}$ gewechselt. Seien $R(Z) := r\Big(1 - \frac{1}{2}(Z + Z^{-1})\Big)$ und $E \in \mathbb{N}$ so, dass $SE \leq A + 2(V-1) \leq SE + S - 1$ gilt. Die Koeffizienten $R_{mS} \in \mathcal{R}$ des Laurent–Polynoms $R(Z)$ sind nur für $m \in \{-E, -E+1, \ldots, E\}$ von Null verschieden. Da die Folge der Koeffizienten von $R(Z)$ symmetrisch ist, ergeben sich die Orthogonalitätsbedingungen aus den Polynomen

$$R_0, R_S \ldots, R_{ES} \in \mathcal{R} = \mathbb{Q}[Y_1, \ldots, Y_V] \,.$$

- Seien Parameter $y = (y_1, \ldots, y_V) \in \mathbb{R}^V$ gefunden, die das System $R_0(y) = \cdots = R_{ES}(y) = 0$, $y_V \neq 0$ erfüllen. Mit $q_y(X) \in \mathbb{R}[X]$ sei das Polynom bezeichnet, welches sich ergibt, wenn im Polynom q die Variablen $Y_1, \ldots, Y_V$ durch die Koordinaten von y ersetzt werden. Dann ergibt sich eine symmetrische orthogonale Skalierungsfolge als $a(Z) = SH_S(Z)^A p(Z)$ mit

$$p(Z) = \frac{1}{2}\left(Z^{-M} + Z^{M-(S-1)(A-1)}\right) q_y\Big(1 - \tfrac{1}{2}(Z + Z^{-1})\Big),$$

  wobei $M \in \mathbb{N}$ durch die Ungleichungen $2M - 1 \leq (S-1)(A-1) \leq 2M$ eindeutig bestimmt ist.

- Erfüllen die Koeffizienten von $p(Z)$ die Ungleichung

$$\max_{k=1,\ldots,S} \sum_{n\in\mathbb{Z}} |p_{k+nS}| < S^{A-1},$$

  so gibt es zur Skalierungsfolge $a \in \ell_{\text{fin}}(\mathbb{R})$ eine stetige symmetrische orthogonale Skalierungsfunktion φ mit kompaktem Träger, die eine orthogonale Multiskalenanalyse erzeugt. Je kleiner die linke Seite der Ungleichung ist, desto glatter ist die Skalierungsfunktion.

Abgesehen von der kleinen Anzahl von Parametertripeln, die in Abschnitt 3.5.7 behandelt wurden, ergeben alle anderen Tripel positiv–dimensionale Lösungsmengen. Da unter den unreduzierten Orthogonalitätsbedingungen an die Folge $a \in \ell_{\text{fin}}(\mathbb{R})$ die Identität

$$\sum_{n\in\mathbb{Z}} a_n^2 = S$$

enthalten ist, bilden die orthogonalen Skalierungsfolgen einer bestimmten Länge eine Teilmenge einer Sphäre. Die Multiplikation endlicher Folgen mit Haar–Faktoren ist linear und injektiv, daher ist auch die reelle Lösungsmenge des Systems $R_0(y) = \cdots = R_{ES}(y) = 0$ kompakt. Es kann ein beliebiges quadratisches Polynom als Zielfunktion des Optimierungsproblems vorgegeben werden. Als einfachste Variante bietet sich die Quadratsumme

$$\sum_{n\in\mathbb{Z}} p_n^2$$

der Koeffizienten des Laurent–Polynoms $p(Z)$ an.

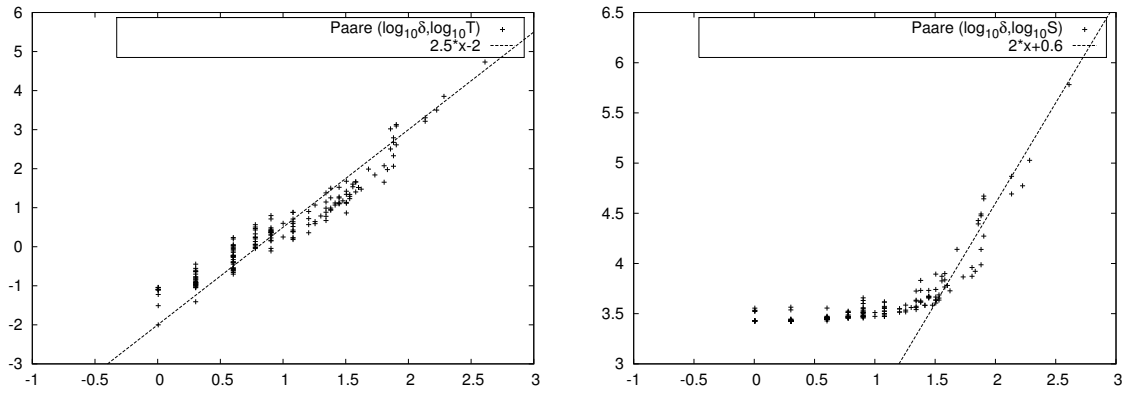


Abbildung 6.2: Darstellung der Paare aus den Logarithmen von geometrischem Grad δ und Laufzeit T bzw. Speicherbedarf S

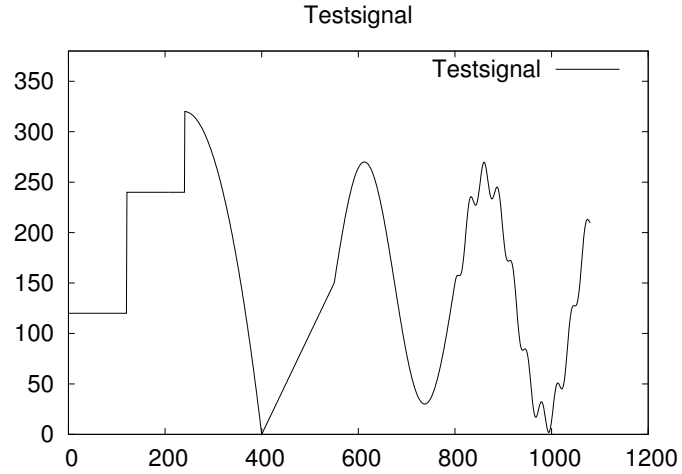


Abbildung 6.3: Das gewählte Testsignal

6.4.2 Allgemeine Bemerkungen zu den berechneten Beispielen

Die Gleichungssysteme, die sich aus der Geometrie der polaren Varietäten zu diesen Vorgaben ergeben, wurden mit dem TERA–Kronecker–Algorithmus gelöst. Trägt man die Rechenzeiten und Speicheranforderungen der berechneten Beispiele zusammen und trägt sie über dem geometrischen Grad ab, so ergeben sich die in Abbildung 6.2 dargestellten Graphen. Es wurde jeweils eine geschätzte Gerade hinzugefügt, die die Punktwolke für große geometrische Grade gut repräsentiert. Somit kann abgelesen werden, dass für diese Beispielklasse sich die Laufzeit proportional zu $\delta^{2.5}$ und der Speicherbedarf proportional zu δ^2 verhalten.

Die Qualität der nach dem Algorithmus der polaren Varietäten bestimmten Lösungen kann auf verschiedene Weise veranschaulicht werden. Zum ersten kann eine Schätzung des Hölder–Index bestimmt werden. Als visueller Gegenpart zu dieser Zahl kann der Graph der Skalierungsfunktion dargestellt werden.

Als dritte Möglichkeit wurde die Eignung zur Datenkompression geprüft. Dazu wurde ein

Testsignal mit Sprüngen und Knickstellen erzeugt, s. Abb. 6.3. Zu jeder der symmetrisch-orthogonalen Skalierungsfolgen wurde eine in gleicher Art symmetrische und orthogonale Wavelet-Filterbank erzeugt. Das Testsignal wurde mit einer 3-stufigen Analyse-Filterbankkaskade transformiert und die entstehenden Koeffizienten der Größe nach sortiert. Eine Datenkompression wird nun dadurch simuliert, dass die kleinsten Koeffizienten in einem vorgegebenen Anteil zu Null gesetzt werden. Nachfolgend wird mit der inversen 3-stufigen Synthese-Filterbankkaskade das reduzierte Signal rekonstruiert.

6.4.3 Beispiele zum Skalenfaktor $S = 3$

Die mit den zur Verfügung stehenden Mitteln berechenbaren Wavelet-Transformationen zum Skalenfaktor $S = 3$ lassen sich wie folgt tabellieren

	V=1	V=2	V=3	V=4	V=5	V=6	V=7
A=1	1	-	4(8)	4(8)	0		
A=2	2	2	-	8(18)	8(18)		
A=3	-	2(4)	2(4)	-	8(40)		
A=4	-	-	4(8)	4(8)	-		
A=5	-	-	-	4(16)	4(16)	-	20(192)
A=6	-	-	-	-	8(32)	8(32)	-
A=7	-	-	-	-	-	8(64)	8(64)

Dabei wurde zu jedem Paar (A, V) die Anzahl der gefundenen reellen Lösungen eingetragen, in Klammern dahinter die Anzahl der komplexen Lösungen, sofern diese größer ist als die Anzahl der reellen Lösungen. Ein Minuszeichen steht für ein unlösbares System.

Als Beispiel für die Struktur der Lösungsmenge wurden in Abbildung 6.4 die Skalierungsfunktionen samt Schätzung ihres Hölder-Index zum Parametertripel $(S, A, V) = (3, 3, 5)$ zusammengetragen. Neben den zwei hutförmigen einmal stetig differenzierbaren Lösungen gibt es noch zwei stetige Lösungen, die die fraktale Struktur der Skalierungsfunktion erkennen lassen. Die weiteren vier Lösungen sind unstetig.

Die einfachste Wavelet-Transformation zu jedem Skalenfaktor $S \in \mathbb{N}_{\geq 2}$ ist die kürzeste mit Approximationsordnung 1. Die Polyphasenmatrix zur Skalierungsfolge ist ein Spaltenvektor der Länge S, dessen Komponenten sämtlich 1 sind. Dieser Vektor kann einfach zu einer reellen orthogonalen Matrix ergänzt werden, hier wurde die Fortsetzung mit der Matrix der diskreten Kosinustransformation (DCT) gewählt. Für $S = 2$ ergibt sich aus dieser Konstruktion die klassische Haar-Wavelet-Transformation. Deshalb bietet sich *Haar-DCT-Wavelet-Transformation* als allgemeine Bezeichnung an. Im Fall $S = 3$ ergeben sich die in Abbildung 6.5 dargestellten, stückweise konstanten Skalierungs- und Waveletfunktionen.

Alle weiteren, optimierten Wavelet-Transformationen bauen auf der DCT auf, d.h. die DCT–Matrix wird zur Vervollständigung der Faktorisierung des Polyphasenvektors der Skalierungsfolge verwendet. Damit wird der Rechenaufwand zur Realisierung einer solchen Transformation immer über dem für das reine Haar–DCT–Wavelet notwendigen liegen. Es ist zu erwar-

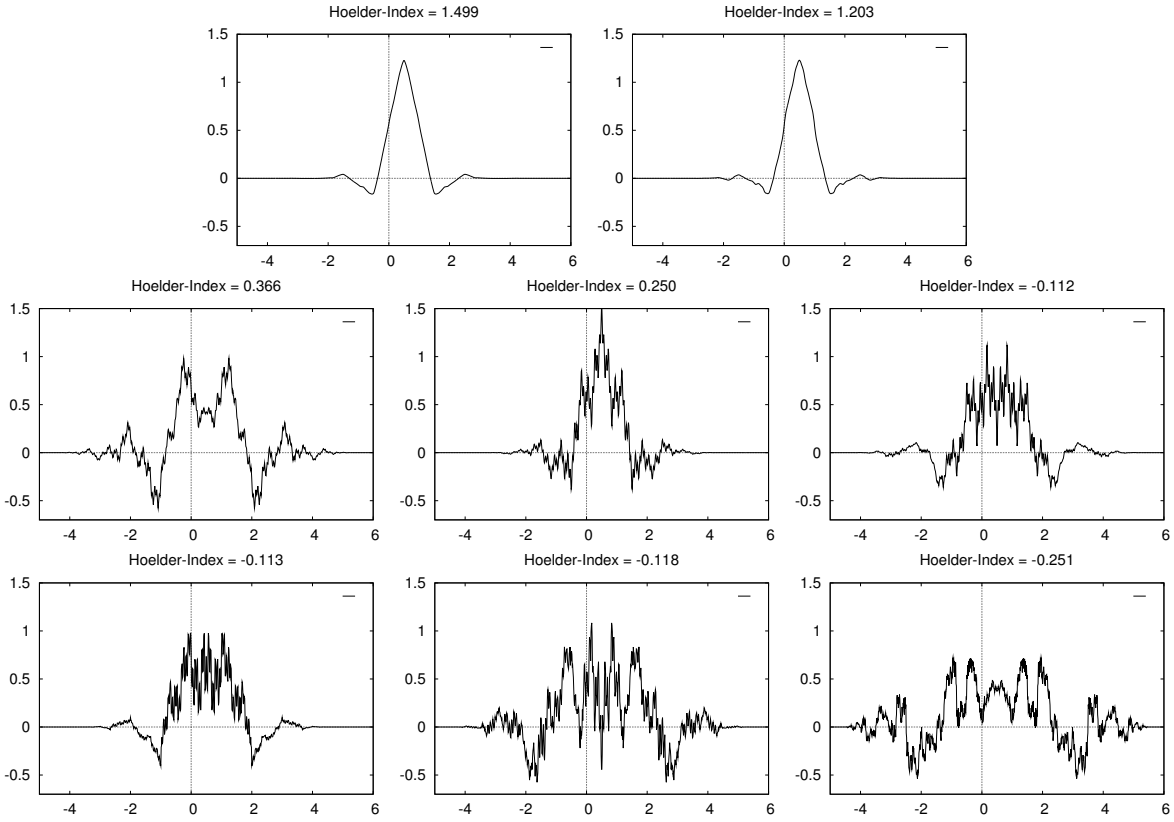


Abbildung 6.4: Die Skalierungsfunktionen zu den kritischen Punkten des Parametertripels $(S, A, V) = (3, 3, 5)$

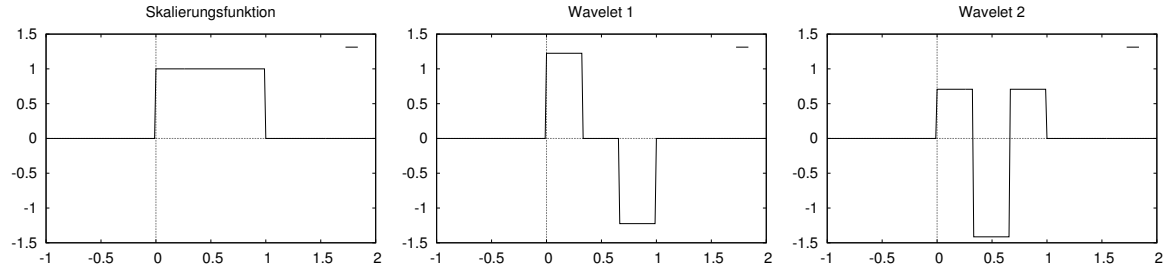


Abbildung 6.5: DCT–Haar–Wavelets zum Skalenfaktor $S = 3$.

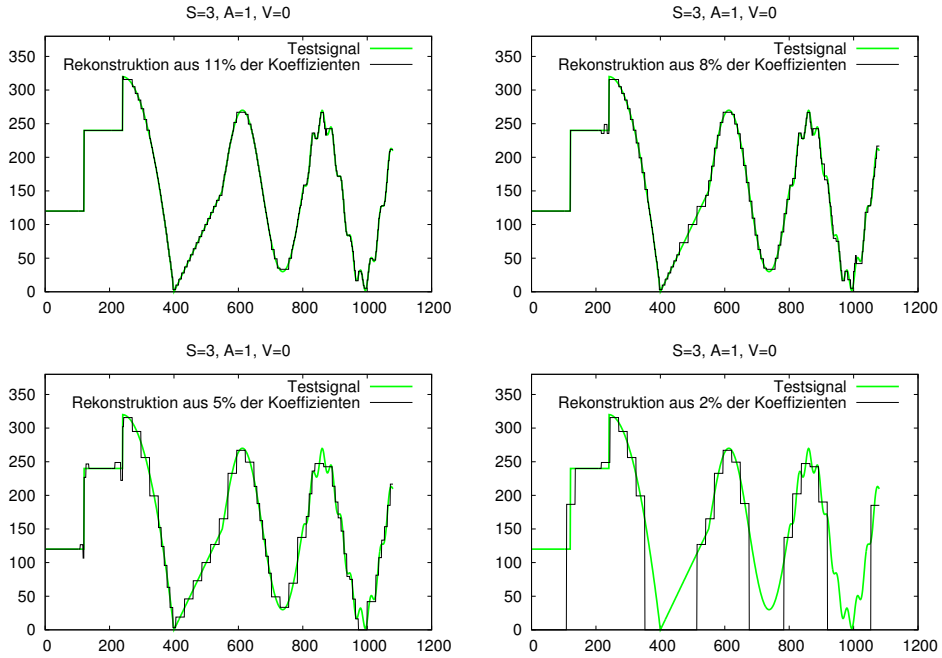


Abbildung 6.6: Kompressionstest für die Haar-DCT-Wavelet-Transformation zum Skalenfaktor 3. Das Testsignal ist zum Vergleich unterlegt.

ten, dass der vergrößerte Rechenaufwand sich in besseren analytischen wie Kompressionseigenschaften niederschlägt.

In den Abbildungen 6.6 und 6.8 wurde zur Veranschaulichung dieses Effekts die testweise Kompression für das Haar–DCT–Wavelet und eine optimierte Wavelet-Transformation zum Parametertripel $(S, A, V) = (3, 5, 4)$ dargestellt. Die Skalierungsfunktion der letztgenannten Transformation (s. Abbildung 6.7) ist einmal stetig differenzierbar, für ihre Ableitung erhält man 0.64 als Schätzung des Index der Hölder–Stetigkeit.

Der Einfluss der Glattheit der Wavelet- und Skalierungsfunktionen ist im Vergleich beider Bildreihen deutlich zu erkennen. Ihrer Natur nach erzeugen die Haar–DCT–Wavelets stückweise konstante Approximationen. In der Bildkompression wird dieser Effekt als Blockbildung bezeichnet. Dafür entstehen an den Sprungstellen des Signals geringere Störungen. Um-

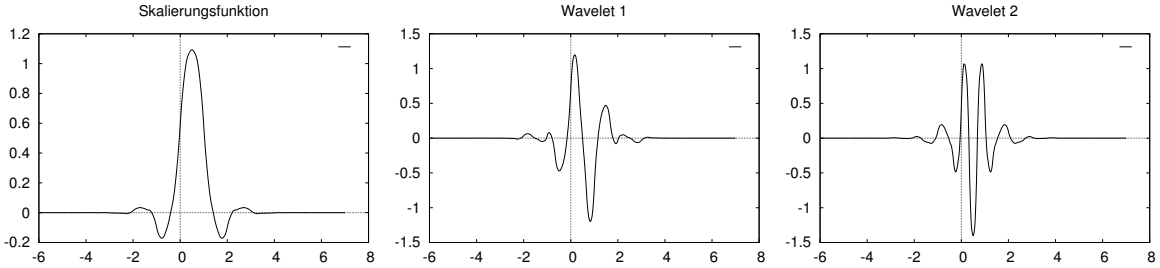


Abbildung 6.7: Optimales Wavelet zum Skalenfaktor $S = 3$, Approximationsordnung $A = 5$, $V = 4$ Freiheitsgrade.

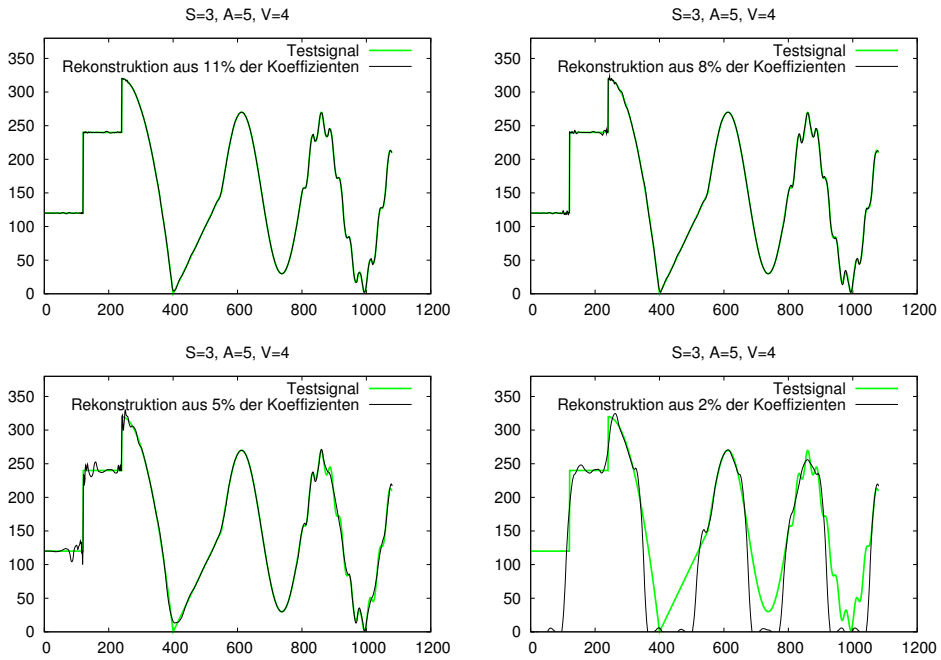


Abbildung 6.8: Kompressionstest für eine optimierte Wavelet-Transformation. Das Testsignal ist zum Vergleich unterlegt.

gekehrt ist für das Beispiel der stetigen Skalierungs- und Wavelet-Funktionen die Approximation im glatten Teil des Testsignals sichtbar besser bei gleicher Kompressionsrate. An den Sprungstellen entstehen jedoch ausgedehntere Bereiche mit Fehlern. Man erkennt die für die Wavelet-Kompression typische Wellenbildung. In beiden Fällen ist ersichtlich, dass 2% der Koeffizienten nicht mehr zur vollständigen Rekonstruktion des Tiefpassanteils ausreichen.

In Abbildung 6.9 ist die Entwicklung des relativen Fehlers in der Rekonstruktion des Testsignals aus verschiedenen Anteilen von Koeffizienten für verschiedene Wavelet-Transformationen zusammengestellt. Die Wavelet-Transformationen entsprechen den besten Lösungen mit einer mindestens einmal differenzierbaren Skalierungsfunktion. Die senkrechte Achse ist als Güte zu lesen, ein Wert von $30\,dB$ entspricht einem relativen Fehler von 10^{-3}, ein Wert von $80\,dB$ analog einem Fehler von 10^{-8}. Die erste Zahl von $30\,dB$ wird im Allgemeinen als Wahrnehmungsschranke angenommen, kleinere relative Fehler wie z.B. $80\,dB$ sind mit menschlichen Sinnen nicht mehr wahrnehmbar. Die Wahrnehmungsschranke wird für die glatten Wavelets schon bei etwa 20% der Koeffizienten erreicht, während für das Haar–DCT–Wavelet 40% der Koeffizienten notwendig sind.

6.4.4 Beispiele zum Skalenfaktor $S = 4$

Die mit den zur Verfügung stehenden Mitteln berechenbaren Wavelet-Transformationen zum Skalenfaktor $S = 4$ sind in nachfolgender Tabelle angegeben.

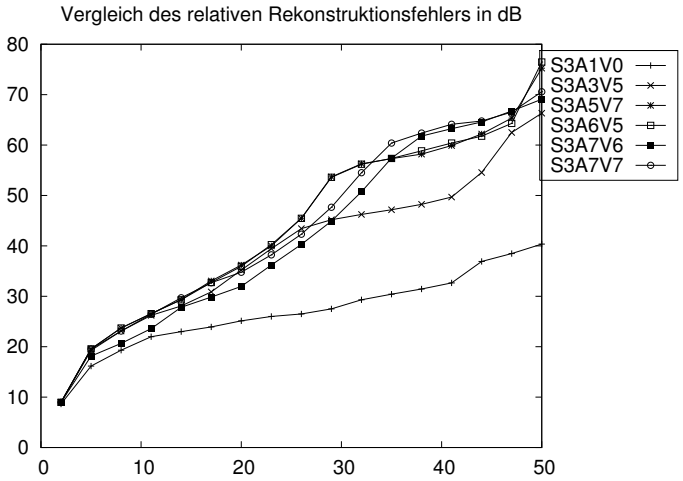


Abbildung 6.9: Entwicklung des relativen Fehlers in der Rekonstruktion des Testsignals für verschiedene Wavelet-Transformationen

	V=1	V=2	V=3	V=4	V=5	V=6	V=7
A=1	1	4	3(4)	6(16)	6(16)		
A=2	2	2	6(10)	6(10)	18(42)		
A=3	0(2)	2	6(12)	8(12)	24(54)		
A=4	-	2(4)	2(4)	10(28)	10(28)	24(136)	
A=5	-	-	0(4)	8(32)	10(32)	36(168)	
A=6	-	-	0(8)	2(8)	16(72)	12(72)	44(408)
A=7	-	-	0(8)	0(8)	4(80)	12(80)	

In Abbildung 6.10 ist die Entwicklung des relativen Fehlers in der Rekonstruktion für ausgesuchte Wavelet-Transformationen dargestellt. Der Trend bzgl. der Güte der Rekonstruktion

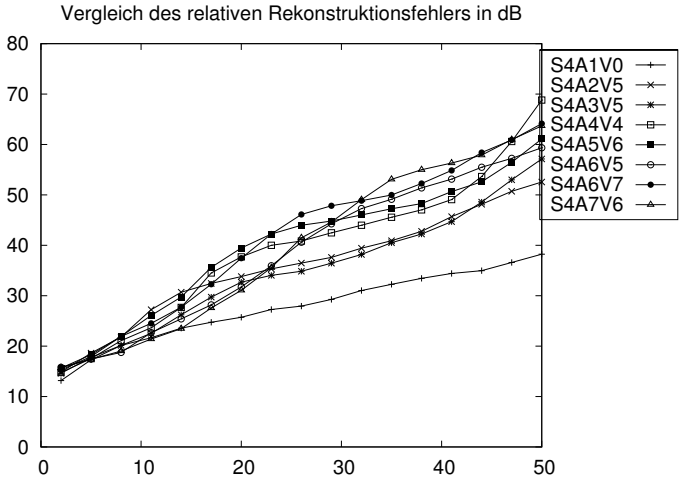


Abbildung 6.10: Entwicklung des relativen Fehlers in der Rekonstruktion des Testsignals für verschiedene Wavelet-Transformationen

Abbildung 6.11: Barb–Testbild aus der Kompressionstestserie, zu finden z.B. unter http://links.uwaterloo.ca/BragZone/

lässt sich auch hier ablesen, die Wahrnehmungsschranke von $30\,dB$ wird hier für die glatten Wavelet-Transformationen wieder unter Benutzung von ca. 20% der Koeffizienten unterschritten. Das Haar–DCT–Wavelet zum Skalenfaktor $S = 4$ erreicht diese Schranke erst bei 30% der Koeffizienten.

Eine weitere Variante zur Visualisierung der mit der Wavelet-Transformation möglichen Konpression besteht in der direkten Anwendung dieser auf Testbilder (wie z.B. in Abbildung 6.11). Analog zum eindimensionalen Testsignal wurde in beiden Dimensionen des Bildes eine dreistufige Filterbankkaskade angewandt. Bei einmaliger Anwendung einer Analyse–Filterbank zum Skalenfaktor $S = 4$ entstehen ein Tiefpassteilbild und 15 Hochpassanteile. Die nächste Stufe der Filterbankkaskade wird nur auf das Tiefpassteilbild angewandt. Es entsteht eine Struktur wie sie im linken Bild von Abbildung 6.12 dargestellt ist. Das rechte Bild dieser Abbildung stellt die Größe der Koeffizienten nach der dreistufigen Filterbankkaskade dar.

Um überhaupt wahrnehmbare Unterschiede zwischen den Rekonstruktionen zu verschiedenen Wavelet-Transformationen darstellen zu können, darf diese aus maximal 10% der nach Größe sortierten Koeffizienten erfolgen. Bei einer Rekonstruktion aus 20% der Koeffizienten gibt es kaum noch wahrnehmbare Unterschiede zum Originalbild. Um den schon erwähnten Unterschied zwischen den Haar–DCT–Wavelets und glatten Wavelets hervorzuheben, ist in Abbildung 6.13 die Rekonstruktion aus 3% dargestellt. Dabei ist links die Rekonstruktion zum Haar–DCT–Wavelet und rechts zum Wavelet mit Parametertripel $(S, A, V) = (4, 6, 7)$ dargestellt. Die Skalierungsfunktion zu diesem Wavelet ist zweimal stetig differenzierbar, ihre

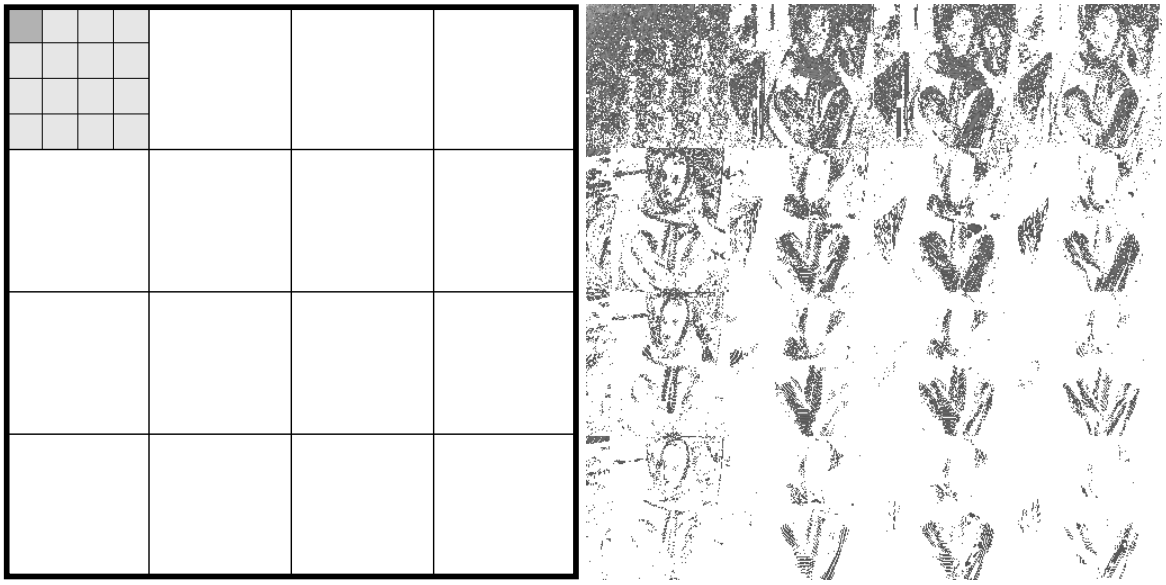


Abbildung 6.12: Links: Schema der Anordnung der Wavelet–Koeffizienten nach einer zweistufigen Wavelet–Filterbank–Kaskade. Rechts: Koeffizienten nach einer dreistufigen Wavelet–Filterbank–Kaskade. Die Graustufe ist vom Logarithmus des Absolutbetrags des Koeffizienten abhängig.

zweite Ableitung hat einen Hölder–Index von 0.77.

Die Rekonstruktion zum Haar–DCT–Wavelet weist die für die JPEG–Kompression typischen Blockartefakte auf. Diese entstehen durch die stückweise konstante Natur der in der Rekonstruktion verwendeten Funktionen. Die Rekonstruktion zum glatten Wavelet besitzt keine solchen Blöcke, hingegen lassen sich wellenförmige Echos an Kanten mit scharfen Kontrasten beobachten.

In Abbildung 6.14 ist zusätzlich noch die Rekonstruktion aus 8% der nach Größe sortierten Koeffizienten dargestellt. Im linken Bild zum Haar–DCT–Wavelet sind die Blöcke klein genug, um kaum noch aufzufallen, jedoch erscheinen schräge Kanten noch stufig. Im rechten Bild zum glatten Wavelet sind auch schräge Kanten glatt, jedoch sind noch leichte Echos an Stellen starker Kontraste zu erkennen.

Abbildung 6.13: Vergleich der Rekonstruktion aus 3% der Koeffizienten der Wavelet-Transformation. Links mittels Haar–DCT–Wavelet, rechts mittels Wavelet zum Parametertripel $(S, A, V) = (4, 6, 7)$

Abbildung 6.14: Vergleich der Rekonstruktion aus 8% der Koeffizienten der Wavelet-Transformation. Links mittels Haar–DCT–Wavelet, rechts mittels Wavelet zum Parametertripel $(S, A, V) = (4, 6, 7)$

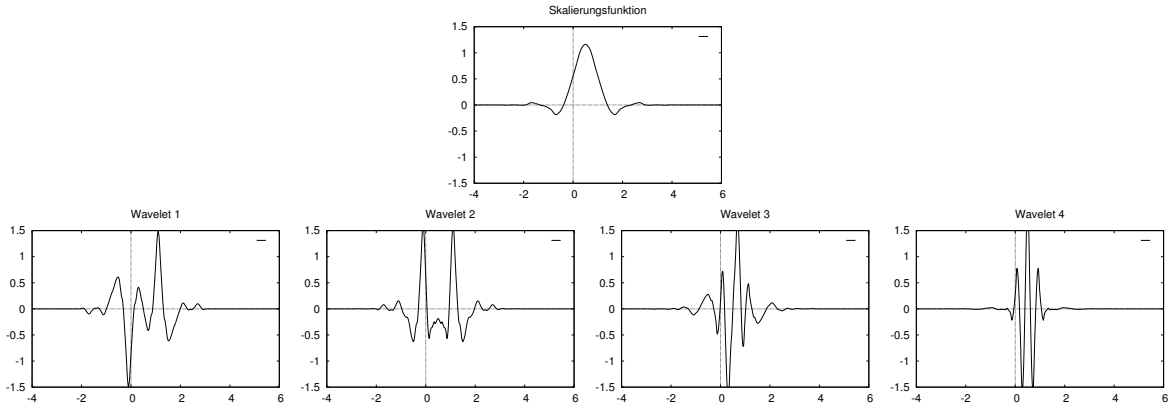


Abbildung 6.15: Skalierungsfunktion und Wavelets für das Parametertripel $(S, A, V) = (5, 4, 5)$

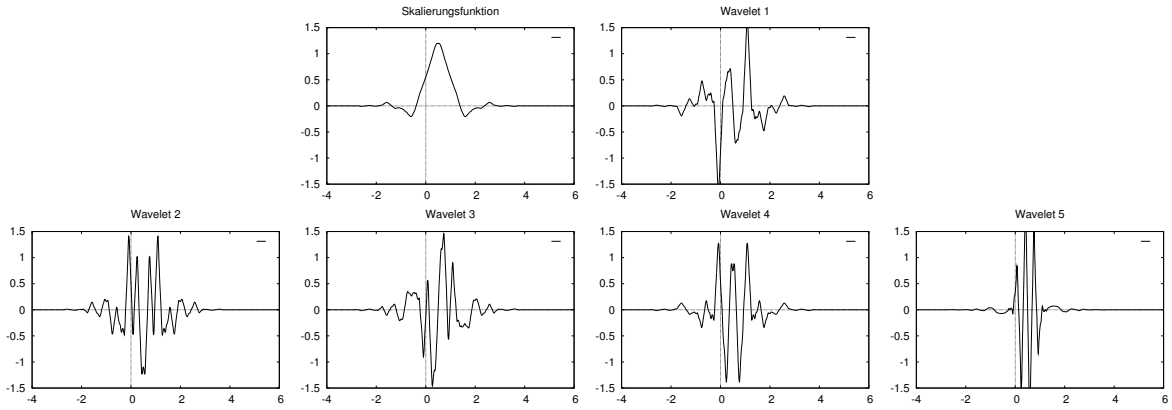


Abbildung 6.16: Skalierungsfunktion und Wavelets für das Parametertripel $(S, A, V) = (6, 4, 5)$

6.4.5 Zusammenfassung der weiteren Rechenergebnisse

Die mit den zur Verfügung stehenden Mitteln berechenbaren Wavelet-Transformationen zum Skalenfaktor $S = 5$ sind in der nachfolgenden Tabelle zusammengefasst. Das glatteste gefundene Wavelet ist das zum Parametertripel $(S, A, V) = (5, 4, 5)$, seine Funktionen sind in Abbildung 6.15 dargestellt. Die Skalierungsfunktion ist einmal stetig differenzierbar, ihre Ableitung ist Hölder–stetig mit Index 0.68.

	V=1	V=2	V=3	V=4	V=5	V=6	V=7
A=1	1	4	-	10(18)	8(26)		
A=2	2	4	4	12(22)	-		
A=3	0(2)	-	4(12)	10(22)	14(22)		
A=4	0(2)	0(2)	2(12)	-	18(68)	24(136)	
A=5	-	-	0(12)	0(12)	10(76)	-	
A=6	-	-	-	0(32)	0(76)	10(76)	
A=7	-	-	0(4)	0(32)	-		

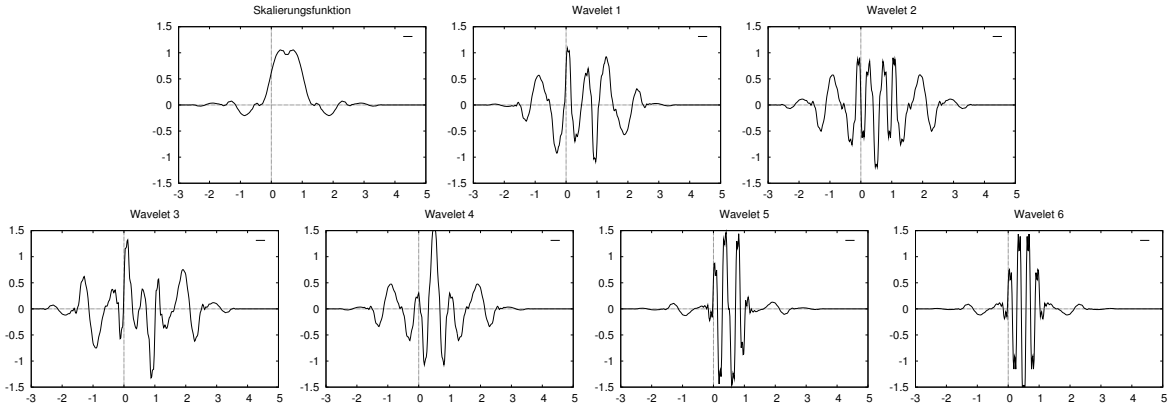


Abbildung 6.17: Skalierungsfunktion und Wavelets für das Parametertripel $(S, A, V) = (7, 4, 5)$

Die berechneten Wavelet-Transformationen zum Skalenfaktor $S = 6$ sind in der nachfolgenden Tabelle zusammengefasst. Die dabei gefundene glatteste Lösung ist die Wavelet-Transformation zum Parametertripel $(S, A, V) = (6, 4, 5)$, s. Abbildung 6.16. Die zugehörige Skalierungsfunktion ist einfach stetig differenzierbar und hat eine Ableitung mit Hölder–Index 0.56.

	V=1	V=2	V=3	V=4	V=5	V=6	V=7
A=1	1	4	6	4(6)	14(30)		
A=2	2	4	2(4)	10(22)	16(28)		
A=3	0(2)	0(4)	4	8(24)	16(34)		
A=4	0(2)	0(2)	0(12)	6(22)	10(22)		
A=5	-	-	0(12)	0(24)	0(24)		
A=6	-	-	0(12)	0(12)	0(76)		
A=7	-	-	0(12)	0(12)	0(80)		

Die berechneten Wavelet-Transformationen zum Skalenfaktor $S = 7$ sind in der nachfolgenden Tabelle zusammengefasst. Die dabei gefundene glatteste Lösung ist die Wavelet-Transformation zum Parametertripel $(S, A, V) = (7, 4, 5)$, s. Abbildung 6.17. Die zugehörige Skalierungsfunktion ist einfach stetig differenzierbar und hat eine Ableitung mit Hölder–Index 0.32.

	V=1	V=2	V=3	V=4	V=5
A=1	1	4	6	-	12(26)
A=2	0(2)	4	6	4(6)	16(34)
A=3	0(2)	0(4)	-	8(20)	12(38)
A=4	0(2)	0(4)	0(4)	0(24)	6(38)

Die mit den zur Verfügung stehenden Mitteln berechenbaren Wavelet-Transformationen zum Skalenfaktor $S = 8$ sind in der nachfolgenden Tabelle zusammengefasst. Die dabei bestimmte Wavelet mit der glattesten Skalierungsfunktion ist in Abbildung 6.18 dargestellt. Die Skalierungsfunktion ist einmal stetig differenzierbar, die Ableitung ist Hölder–stetig mit Index 0.71.

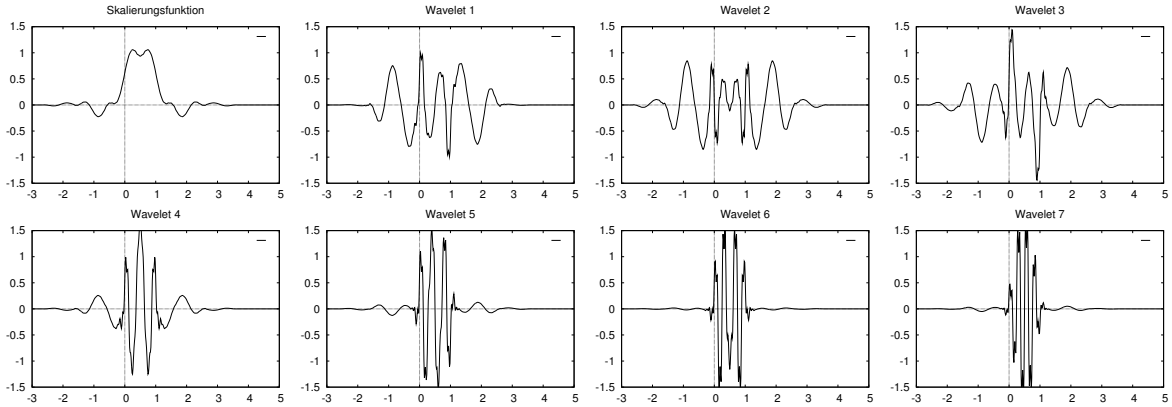


Abbildung 6.18: Skalierungsfunktion und Wavelets für das Parametertripel $(S, A, V) = (8, 4, 6)$

	V=1	V=2	V=3	V=4	V=5	V=6
A=1	1	4	6	8	4(8)	
A=2	0(2)	4	6	2(6)	16(34)	
A=3	0(2)	0(4)	0(6)	4(6)	14(36)	
A=4	0(2)	0(4)	0(4)	0(24)	0(38)	14(48)

Anhang A

Normierte Folgen– und Funktionenräume

A.1 Folgenräume

Mit $\ell(\mathbb{Z}) := \ell(\mathbb{Z}, \mathbb{C})$ bezeichnen wir den Raum aller Folgen $a : \mathbb{Z} \to \mathbb{C}$, die wir als $a = \{a_n\}_{n\in\mathbb{Z}}$ notieren. Dieser ist ein unendlichdimensionaler $\mathbb{C}$–Vektorraum. Eine Folge $a = \{a_n\}_{n\in\mathbb{Z}}$ nennen wir *endlich*, wenn die Menge $\{n \in \mathbb{Z} : \ a_n \neq 0\}$ ihrer nichtverschwindenden Glieder endlich ist. Sei mit $\ell_{\text{fin}}(\mathbb{Z})$ die Teilmenge der endlichen Folgen von $\ell(\mathbb{Z})$ bezeichnet.

Dann können auf $\ell_{\text{fin}}(\mathbb{Z})$ die Normen

$$\|a\|_p := \left(\sum_{n\in\mathbb{Z}} |a_n|^p\right)^{\frac{1}{p}}, \ 1 \leq p < \infty, \quad \text{und} \quad \|a\|_\infty := \sup_{n\in\mathbb{Z}} |a_n|$$

für jedes $a \in \ell_{\text{fin}}(\mathbb{Z})$ definiert werden. Mit $\ell_p(\mathbb{Z})$ bezeichnen wir den Abschluss von $\ell_{\text{fin}}(\mathbb{Z})$ in $\ell(\mathbb{Z})$ bzgl. der Norm $\|\cdot\|_p$. D.h. $\ell_p(\mathbb{Z})$ besteht aus den Folgen in $\ell(\mathbb{Z})$, für welche die in der Definition der Norm vorkommende Reihe konvergiert. Auf $\ell_2(\mathbb{Z})$ läßt sich ein hermitisches Skalarprodukt definieren,

$$\langle\, a,\, b\,\rangle_{\ell_2} := \sum_{n\in\mathbb{Z}} a_n\overline{b_n}, \qquad \forall a, b \in \ell_2(\mathbb{Z})\,,$$

welches mit der Norm auf $\ell_2(\mathbb{Z})$ verträglich ist, d.h. $\langle\, a,\, a\,\rangle_{\ell_2} = \|a\|^2_{\ell_2}$. Mit diesem Skalarprodukt wird $\ell_2(\mathbb{Z})$ zum Hilbert–Raum.

Es ist $\ell_1(\mathbb{Z}) \subset \ell_p(\mathbb{Z}) \subset \ell_\infty(\mathbb{Z})$, und diese Inklusionen sind für $1 < p < \infty$ strikt. Jedoch können Folgen aus $\ell_p(\mathbb{Z})$ durch Folgen aus $\ell_1(\mathbb{Z})$ beliebig genau approximiert werden, was der folgende Satz besagt.

Satz A.1.1 *Seien $b_n : [0,1] \to [0,1]$, $n \in \mathbb{Z}$, stetige Funktionen mit $\lim_{\lambda\to 0} b_n(\lambda) = 1$ für jedes $n \in \mathbb{Z}$ und $\{b_n(\lambda)\}_{n\in\mathbb{Z}} \in \ell_1(\mathbb{Z})$ für jedes $\lambda \in (0,1]$.*

Dann ist für jedes $a \in \ell_p(\mathbb{Z})$ bei $p \in (1,\infty)$ die Familie der Folgen $a_\lambda := \{a_n b_n(\lambda)\}_{n\in\mathbb{Z}}$, $\lambda \in (0,1]$, in $\ell_1(\mathbb{Z})$ enthalten und $\lim_{\lambda\to 0} \|a_\lambda - a\|_p = 0$, d.h. a_λ konvergiert in $\ell_p(\mathbb{Z})$ gegen a.

Beweis: Da $\sum_{n\in\mathbb{Z}} |a_n|^p < \infty$ gilt, fällt die Folge a im Unendlichen gegen Null ab, $\lim_{|n|\to\infty} a_n = 0$, und hat daher ein Maximum. Somit gilt

$$\sum_{n\in\mathbb{Z}} |a_n b_n(\lambda)| \leq \sum_{n\in\mathbb{Z}} |b_n(\lambda)| \ \max_{n\in\mathbb{Z}} |a_n| < \infty$$

und damit ist $a_\lambda \in \ell_1(\mathbb{Z})$ für jedes $\lambda \in (0,1]$.

Für den Konvergenznachweis sei $\|a\|_p \neq 0$ vorausgesetzt, andernfalls ist die Konvergenzaussage trivial. Sei ein $\varepsilon > 0$ vorgegeben. Dann gibt es ein $N \in \mathbb{N}$ mit $\sum_{n\in\mathbb{Z}:\,|n|>N} |a_n|^p < \varepsilon^p$. Weiterhin kann eine Schranke $\delta > 0$ gefunden werden, so dass für jedes $\lambda \in [0,\delta)$ für die endlich vielen Funktionen b_n, $n = -N,\dots,N$, gemeinsam gilt $|1 - b_n(\lambda)| \leq \frac{1}{\|a\|_p}\varepsilon$. Somit ist für $0 \leq \lambda < \delta$

$$\|a_\lambda - a\|_p^p = \sum_{n\in\mathbb{Z}} |b_n(\lambda)a_n - a_n| \leq \sum_{|n|\leq N} |a_n|\,\|a\|_p^{-p}\varepsilon^p + \sum_{|n|>N} |a_n|^p < 2\varepsilon^p\,,$$

also $\|a_\lambda - a\|_p < 2\varepsilon$. Da $\varepsilon > 0$ beliebig gewählt werden kann, ist die Konvergenz gezeigt. □

A.2 Funktionenräume

A.2.1 Räume stetiger Funktionen

Mit $C(\mathbb{R}) := C(\mathbb{R},\mathbb{C})$ wird der Raum der stetigen komplexwertigen Funktionen bezeichnet. Mit $C_b(\mathbb{R})$ sei die Teilmenge der beschränkten stetigen Funktionen bezeichnet. $C_0(\mathbb{R})$ bezeichnet den Teilraum derjenigen stetigen Funktionen, welche im Unendlichen gegen Null abfallen, d.h. für jedes $f \in C_0(\mathbb{R})$ und $\varepsilon > 0$ gibt es einen Abstand $R > 0$, so dass $|f(x)| < \varepsilon$ für alle $x \in \mathbb{R}$ mit $|x| > R$ gilt. Mit $C_c(\mathbb{R})$ bezeichnen wir den wiederum in diesem enthaltenen Teilraum der stetigen Funktionen mit kompaktem Träger, d.h. für jedes $f \in C_c(\mathbb{R})$ gibt es einen Abstand $R > 0$, so dass $f(x) = 0$ für alle $x \in \mathbb{R}$ mit $|x| > R$ gilt. Als Träger $\operatorname{supp} f$ einer Funktion $f : \mathbb{R} \to \mathbb{C}$ bezeichnen wir den topologischen Abschluss der Menge $\{x \in \mathbb{R} :\ f(x) \neq 0\}$.

A.2.2 Räume messbarer Funktionen

Mit $L^p(\mathbb{R}) := \mathcal{L}^p(\mathbb{R},\mathbb{C})$, $1 \leq p < \infty$, bezeichnen wir den Raum der messbaren Funktionen $f : \mathbb{R} \to \mathbb{C}$, für welche das Lebesgue–Integral $\int_{\mathbb{R}} |f(x)|^p \,\mathrm{d}x$ existiert und endlich ist. Genauer gesagt, besteht dieser Raum aus Äquivalenzklassen von Funktionen, wobei zwei Funktionen als äquivalent angesehen werden, wenn sie sich nur auf einer Menge vom Maß Null unterscheiden, d.h. $\int_{\mathbb{R}} |f(x) - g(x)|^p \,\mathrm{d}x = 0$ gilt. Auf diesem Raum ist $\|f\|_p := \left(\int_{\mathbb{R}} |f(x)|^p \,\mathrm{d}x\right)^{\frac{1}{p}}$ eine Norm, und $L^p(\mathbb{R})$ ist mit dieser Norm vollständig, d.h. ein Banach–Raum.

Ist $I = [a,b] \subset \mathbb{R}$ ein beschränktes Intervall, $-\infty < a < b < \infty$, so bezeichnen wir mit $L^p(I)$ denjenigen Teilraum von $L^p(\mathbb{R})$, der diejenigen Funktionen enthält, die außerhalb des Intervalls I den Wert Null annehmen. Wir werden im wesentlichen nur mit den Räumen $L^1(\mathbb{R})$ und $L^2(\mathbb{R})$ arbeiten und die folgenden Grundeigenschaften dieser Funktionenräume benutzen:

- Definieren wir die Translationsabbildung $\mathcal{T}_s : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ für jedes $s \in \mathbb{R}$ durch $\mathcal{T}_s(f)(x) := f(x-s)$, so erhält diese die Norm in $L^p(\mathbb{R})$. Fixieren wir ein $f \in L^p(\mathbb{R})$, so ist die daraus abgeleitete Abbildung $\mathcal{T}_f : \mathbb{R} \to L^p(\mathbb{R})$, $s \mapsto \mathcal{T}_s(f)$, stetig, insbesondere ist die reelle Abbildung $s \mapsto \|\mathcal{T}_s(f) - f\|_p$ stetig.

- Auf dem Raum $L^2(\mathbb{R})$ wird ein hermitesches Skalarprodukt definiert durch

$$\langle f, g \rangle_{L^2(\mathbb{R})} := \int_{\mathbb{R}} f(x)\overline{g(x)}\,\mathrm{d}x, \qquad f,g \in L^2(\mathbb{R})\,.$$

  Für dieses gilt die *Cauchy–Schwarzsche Ungleichung*, d.h. für beliebige $f,g \in L^2(\mathbb{R})$ ist

$$\langle f, g \rangle_{L^2(\mathbb{R})} \leq \|f\|_{L^2(\mathbb{R})}\,\|g\|_{L^2(\mathbb{R})}\,.$$

  Gleichheit gilt nur bei $f = g$. Mit diesem Skalarprodukt wird $L^2(\mathbb{R})$ zum Hilbert–Raum. Diese Ungleichung schätzt das Skalarprodukt durch die Normen der Faktoren ab.

- Es gilt das *Landau–Resonanztheorem* (s. [Off9X]), nämlich

$$\|f\|_{L^2(\mathbb{R})} = \sup_{g\in L^2(\mathbb{R}):\,\|g\|\leq 1} \left|\langle f, g \rangle_{L^2(\mathbb{R})}\right|$$

  für jedes $f \in L^2(\mathbb{R})$. Dabei wird das Supremum für $g(x) = \frac{1}{\|f\|}f(x)$ angenommen. Dieser Satz kann für beliebige L^p–Räume auf der Basis der Hölder–Ungleichung formuliert werden.

- Es gilt der *Satz von Lebesgue zur dominierten Konvergenz*. Seien $g, f_1, f_2, \cdots \in L^1(\mathbb{R})$ integrable Funktionen, so dass die Funktionenfolge $\{f_n\}_{n\in\mathbb{Z}}$ fast überall punktweise konvergiert und die Folgen $\{|f_n(x)|\}_{n\in\mathbb{Z}}$ fast überall durch $g(x)$ nach oben beschränkt sind. Dann kann die Integration mit der Grenzwertbildung vertauscht werden,

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,\mathrm{d}x = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)\,\mathrm{d}x\,.$$

A.2.3 Faltung und Approximation der Eins

Der Raum $C_c(\mathbb{R})$ der stetigen Funktionen mit kompaktem Träger ist in jedem der Räume $L^p(\mathbb{R})$, $1 \leq p < \infty$, enthalten; er ist sogar dicht in diesen. Weiterhin ist mit $\varphi \in C_c(\mathbb{R})$ und $f \in L^p(\mathbb{R})$ auch deren Produkt $\varphi f \in L^p(\mathbb{R})$. Damit kann die Faltung einer Funktion $\varphi \in C_c(\mathbb{R})$ mit einer Funktion $f \in L^p(\mathbb{R})$ definiert werden als $\varphi * f : \mathbb{R} \to \mathbb{C}$,

$$x \mapsto (\varphi * f)(x) := \int_{\mathbb{R}} \varphi(t) f(x-t)\,\mathrm{d}t\,,$$

und die Funktion $\varphi * f$ ist stetig. Mit Hilfe des Landau–Resonanztheorems überzeugt man sich leicht, dass die Faltung auch für alle $(f,\varphi) \in L^2(\mathbb{R}) \times L^1(\mathbb{R})$ definiert ist und $\|f * \varphi\|_{L^2(\mathbb{R})} \leq \|f\|_{L^2(\mathbb{R})}\|\varphi\|_{L^1(\mathbb{R})}$ gilt.

Definition A.2.1 ([Off9X]) *Eine Funktionenfamilie* $\{\varphi_s : s \in (0,1]\} \subset L^1(\mathbb{R})$ *wird* Approximation der Eins *(d.h. der Einheit der Faltung) genannt, wenn*

i) die Funktionen φ_s fast überall nur nichtnegative Werte annehmen,

ii) $\|\varphi_s\|_{L^1(\mathbb{R})} = \int_{\mathbb{R}} \varphi_s(x)\,\mathrm{d}x = 1$ für jedes $s \in (0,1]$ ist und

iii) für jedes $\delta > 0$ gilt $\lim_{s\to 0} \int_{|x|>\delta} \varphi_s(x)\,\mathrm{d}x = 0$.

Satz A.2.2 *Sei die Familie* $\{\varphi_s : s \in (0,1]\} \subset L^1(\mathbb{R})$ *eine Approximation der Eins. Dann gelten folgende Konvergenzaussagen:*

i) Ist f *messbar, beschränkt und im Nullpunkt stetig, so ist* $\lim_{s\to 0} \int_{\mathbb{R}} \varphi_s(x) f(x)\,\mathrm{d}x = f(0)$.

ii) Ist $f \in L^1(\mathbb{R})$, *so ist für jedes* $s \in (0,1]$ *auch* $\varphi_s * f \in L^1(\mathbb{R})$ *und die Familie der Faltungsprodukte konvergiert in* $L^1(\mathbb{R})$, $\lim_{s\to 0} \varphi_s * f = f$.

iii) Ist $f \in L^2(\mathbb{R})$, *so ist für jedes* $s \in (0,1]$ *auch* $\varphi_s * f \in L^2(\mathbb{R})$ *und die Familie der Faltungsprodukte konvergiert in* $L^2(\mathbb{R})$, $\lim_{s\to 0} \varphi_s * f = f$.

Beweis:

zu i) Sei $M > 0$ eine Schranke von f. Dann gibt es für jedes $\varepsilon > 0$ ein $\delta > 0$, so dass $|f(x) - f(0)| < \varepsilon$ für $|x| < \delta$ gilt, sowie

$$\begin{aligned}\left|\int_{\mathbb{R}} \varphi_s(x) f(x)\,\mathrm{d}x - f(0)\right| &= \left|\int_{\mathbb{R}} \varphi_s(x)(f(x) - f(0))\,\mathrm{d}x\right| \\ &\le \varepsilon \int_{|x|<\delta} \varphi_s(x)\,\mathrm{d}x + 2M \int_{|x|\ge\delta} \varphi_s(x)\,\mathrm{d}x\,.\end{aligned}$$

Im Grenzwert bei $s \to 0$ gilt also $\left|\int_{\mathbb{R}} \varphi_s(x) f(x)\,\mathrm{d}x - f(0)\right| \le \varepsilon$. Da $\varepsilon > 0$ beliebig gewählt war, folgt die Behauptung.

zu ii) Die Differenz $\varphi_s * f - f$ wird analog zu i) umgeformt, um die Abschätzung

$$\|\varphi_s * f - f\|_{L^1(\mathbb{R})} \le \int_{\mathbb{R}^2} \varphi_s(t)\,|f(x-t) - f(x)|\,\mathrm{d}t\,\mathrm{d}x \le \int_{\mathbb{R}} \varphi_s(t)\,\|\mathcal{T}_t f - f\|_{L^1(\mathbb{R})}\,\mathrm{d}t$$

zu erhalten. Die Funktion $\|\mathcal{T}_t f - f\|$ ist stetig und beschränkt, nach i) konvergiert das Integral für $s \to 0$ gegen $\|\mathcal{T}_0 f - f\| = 0$.

zu iii) Hier können wir auf das Landau–Resonanztheorem zurückgreifen. Sei $g \in L^2(\mathbb{R})$ eine beliebige Funktion mit $\|g\| \le 1$. Dann folgt mit der Cauchy–Schwarzschen Ungleichung

$$\begin{aligned}\left|\langle\, \varphi_s * f - f,\, g\,\rangle_{L^2(\mathbb{R})}\right| &\le \int_{\mathbb{R}^2} \varphi_s(t) |f(x-t) - f(x)|\,|g(x)|\,\mathrm{d}t\,\mathrm{d}x \\ &\le \int_{\mathbb{R}} \varphi_s(t)\,\|\mathcal{T}_t f - f\|_{L^2(\mathbb{R})}\,\|g\|_{L^2(\mathbb{R})}\,\mathrm{d}t\end{aligned}$$

Aus demselben Grund wie in ii) konvergiert dieser Ausdruck für $s \to 0$ gegen 0, und damit auch die Norm der Differenz $\|\varphi_s * f - f\|_{L^2(\mathbb{R})}$.

□

A.2.4 Orthonormalsysteme in Hilbert–Räumen

Wir stellen hier Grundeigenschaften von Orthonormalsystemen in Hilbert–Räumen zusammen, insbesondere zur Approximation von Elementen des Hilbert–Raums durch Linearkombinationen in einem Orthonormalsystem.

Definition A.2.3 *Ein* $\mathbb{C}$*–Hilbert–Raum ist ein komplexer Vektorraum mit hermiteschem Skalarprodukt, der bzgl. der vom Skalarprodukt induzierten Norm vollständig ist.*

Eine abzählbare Teilmenge $\{\mathbf{e}_n\}_{n\in\mathbb{Z}}$ *eines* $\mathbb{C}$*–Hilbert–Raumes* $(\mathcal{H}, \langle\ ,\ \rangle)$ *heißt* Orthonormalsystem, *wenn ihre Elemente die Länge* 1 *haben und paarweise senkrecht zueinander stehen,*

$$\forall m,n\in\mathbb{Z}:\ \langle\, \mathbf{e}_m,\, \mathbf{e}_n\,\rangle_{\mathcal{H}} = \delta_{m,n} := \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

Es ist unmittelbar zu sehen, dass für endliche Linearkombinationen $\mathbf{a} := \sum_{m\in I} a_m\mathbf{e}_m$ und $\mathbf{b} := \sum_{n\in J} b_n\mathbf{e}_n$, mit $I, J \subset \mathbb{Z}$ endlich, das Skalarprodukt sich bestimmt als

$$\langle\, \mathbf{a},\, \mathbf{b}\,\rangle_{\mathcal{H}} = \sum_{n\in I\cap J} a_n\overline{b_n}\,.$$

Die rechte Seite definiert ein Skalarprodukt auf dem Raum der endlichen komplexwertigen Folgen. Diese können in den Hilbert–Raum $\ell_2(\mathbb{Z}) := \ell_2(\mathbb{Z},\mathbb{C})$ aller Folgen $c = \{c_n\}_{n\in\mathbb{Z}} \subset \mathbb{C}$, für welche die Reihe $\|c\|^2_{\ell_2} := \sum_{n\in\mathbb{Z}} |c_n|^2$ konvergiert, eingebettet werden.

Da nun, für eine fixierte Folge $c \in \ell_2(\mathbb{Z},\mathbb{C})$, die Norm endlicher Teilsummen von $\sum_{n\in\mathbb{Z}} c_n\mathbf{e}_n$ beschränkt ist durch

$$\|\sum_{n\in I} c_n\mathbf{e}_n\|^2_{\mathcal{H}} = \sum_{n\in I} |c_n|^2 \leq \|c\|^2_{\ell_2}$$

für jede endliche Teilmenge $I \subset \mathbb{Z}$, konvergiert die Reihe $\mathbf{c} := \sum_{n\in\mathbb{Z}} c_n\mathbf{e}_n$ in $\mathcal{H}$ und es gilt $\|\mathbf{c}\|_{\mathcal{H}} = \|c\|_{\ell_2}$.

Wir können diese Konstruktion systematisieren, indem wir eine Abbildung $\mathcal{E} : \ell_2(\mathbb{Z}) \to \mathcal{H}$ definieren durch

$$c \mapsto \mathcal{E}(c) := \sum_{n\in\mathbb{Z}} c_n\mathbf{e}_n\,.$$

Wir werden die Abbildung $\mathcal{E}$ im Weiteren als Fourier–Operator zum Orthonormalsystem $\{\mathbf{e}_n :\ n\in\mathbb{Z}\}$ bezeichnen. Dieser ist ein linearer und isometrischer Operator. Daher gibt es einen adjungierten Operator $\mathcal{E}^* : \mathcal{H} \to \ell_2(\mathbb{Z})$, der durch die Eigenschaft

$$\forall c\in\ell_2(\mathbb{Z}), \mathbf{a}\in\mathcal{H}:\ \langle\, \mathcal{E}^*(v),\, c\,\rangle_{\ell_2} = \langle\, \mathbf{a},\, \mathcal{E}(c)\,\rangle$$

definiert ist. Indem wir die Definition einsetzen und umformen, erhalten wir $\mathcal{E}^*(\mathbf{a}) = \{\langle\, \mathbf{a},\, \mathbf{e}_n\,\rangle\}_{n\in\mathbb{Z}}$. Die Glieder dieser Folge werden *Fourier–Koeffizienten* bzgl. des Orthonormalsystems $\{\mathbf{e}_n :\ n\in\mathbb{Z}\}$ genannt.

Lemma A.2.4 *Die Operatornorm des adjungierten Operators* A^* *eines beschränkten linearen Operators* $A : \mathcal{H} \to \mathcal{H}'$ *zwischen zwei* $\mathbb{C}$*–Hilbert–Räumen stimmt mit dessen Operatornorm überein.*

Beweis: Es gilt für jedes $\mathbf{b} \in \mathcal{H}'$ mit $A^*\mathbf{b} \neq 0$

$$\begin{aligned}\|A^*\mathbf{b}\|_{\mathcal{H}} &= \left\langle\, A^*\mathbf{b},\, \tfrac{A^*\mathbf{b}}{\|A^*\mathbf{b}\|}\,\right\rangle_{\mathcal{H}} \leq \sup_{\|\mathbf{a}\|\leq 1} |\,\langle\, A^*\mathbf{b},\, \mathbf{a}\,\rangle_{\mathcal{H}}\,| \\ &= \sup_{\|\mathbf{a}\|\leq 1} |\,\langle\, \mathbf{b},\, A\mathbf{a}\,\rangle_{\mathcal{H}'}\,| \leq \sup_{\|\mathbf{a}\|\leq 1} \|\mathbf{b}\|_{\mathcal{H}'}\,\|A\mathbf{a}\|_{\mathcal{H}'} \leq \|A\|\,\|\mathbf{b}\|\,.\end{aligned}$$

Daher gilt $\|A^*\| \leq \|A\|$. Da für beschränkte Operatoren im Hilbert–Raum $(A^*)^* = A$ gilt, gilt auch $\|A\| \leq \|A^*\|$ und damit die Gleichheit beider Normen. □

Daraus und aus $\|\mathcal{E}\| = 1$ ergibt sich unmittelbar $\|\mathcal{E}^*\| = 1$ und damit die Bessel–Ungleichung $\|\mathcal{E}^*(\mathbf{a})\|_{\ell_2} \leq \|\mathbf{a}\|_{\mathcal{H}}$:

Satz A.2.5 *Sei ein Orthonormalsystem $\{\mathbf{e}_n\}_{n\in\mathbb{Z}}$ eines $\mathbb{C}$–Hilbert–Raumes $\mathcal{H}$ gegeben. Dann gilt für jedes Element $\mathbf{a} \in \mathcal{H}$ die Ungleichung*

$$\sum_{n\in\mathbb{Z}} |\langle \mathbf{a}, \mathbf{e}_n \rangle|^2 \leq \|\mathbf{a}\|_{\mathcal{H}}^2 ,$$

und die Gleichheit gilt genau für die Elemente **a**, *die mit ihrer* Fourier–Reihe

$$\mathcal{E}(\mathcal{E}^*(\mathbf{a})) = \sum_{n\in\mathbb{Z}} \langle \mathbf{a}, \mathbf{e}_n \rangle \mathbf{e}_n$$

übereinstimmen.

Beweis: Die Fourier–Reihe von $\mathbf{a} \in \mathcal{H}$ ist immer die Bestapproximation im vom Orthonormalsystem aufgespannten Unterraum, denn es gilt für eine beliebige Folge $c \in \ell_2(\mathbb{Z})$

$$\begin{aligned}\|\mathbf{a} - \mathcal{E}(c)\|_{\mathcal{H}}^2 =& \|\mathbf{a}\|_{\mathcal{H}}^2 - 2Re(\langle \mathbf{a}, \mathcal{E}(c) \rangle_{\mathcal{H}}) + \|\mathcal{E}(c)\|_{\mathcal{H}}^2 \\ =& \|\mathbf{a}\|_{\mathcal{H}}^2 - 2Re(\langle \mathcal{E}^*(\mathbf{a}), c \rangle_{\ell_2}) + \|c\|_{\ell_2}^2 \\ =& \|\mathbf{a}\|_{\mathcal{H}}^2 - \|\mathcal{E}^*(\mathbf{a})\|_{\ell_2}^2 + \|\mathcal{E}^*(\mathbf{a}) - c\|_{\ell_2}^2 .\end{aligned}$$

Somit wird der minimale Abstand für $c = \mathcal{E}^*(\mathbf{a})$ angenommen. Da dieser minimale Abstand nichtnegativ ist, gilt $\|\mathcal{E}^*(\mathbf{a})\|_{\ell_2}^2 \leq \|\mathbf{a}\|_{\mathcal{H}}^2$, woraus sich die Bessel–Ungleichung ergibt, sowie Gleichheit genau dann, wenn $\mathbf{a} = \mathcal{E}(c) = \mathcal{E}(\mathcal{E}^*(\mathbf{a}))$ eintritt. □

Für jedes Orthonormalsystem in $\mathcal{H}$ gilt die Identität $\mathcal{E}^* \circ \mathcal{E} = id_{\ell_2(\mathbb{Z})}$. Damit ist der Operator $P := \mathcal{E} \circ \mathcal{E}^*$, der jedem Element aus $\mathcal{H}$ seine Fourier–Reihe zuordnet, ein orthogonaler Projektor in $\mathcal{H}$, d.h. $P = P^*$ und $P^2 = P$. Operatoren mit diesen Eigenschaften nennt man auch *partiell unitär*.

Definition A.2.6 *Ein Orthonormalsystem in einem Hilbert–Raum, in welchem jedes Element des Hilbert–Raums als Fourier–Reihe dargestellt werden kann, heißt* vollständig. *Ein vollständiges Orthonormalsystem nennt man auch* Hilbert–Basis.

A.3 Differenzenoperatoren mit unendlichem Träger

Seien V ein endlichdimensionaler und W ein beliebiger $\mathbb{C}$–Hilbert–Raum. Die bisher definierten Differenzenoperatoren der Form $F := f_{-M}\mathcal{T}^{-M} + \cdots + f_M\mathcal{T}^M$ mit $f_{-M}, \ldots, f_M : V \to W$ linear und beschränkt, welche endliche in endliche Folgen abbilden, können problemlos auf unendliche Folgen ausgedehnt werden, d.h. zu jedem $p \in [0, \infty]$ ist $F : \ell_p(V) \to \ell_p(W)$ ein beschränkter linearer Operator. Denn die Verschiebung erhält die Norm jeder Folge, die da

die f_k als beschränkt vorausgesetzt sind, verändert sich die Norm einer Folge bei simultaner Anwendung eines f_k höchstens um einen Faktor, der eine Schranke von f_k ist.

Statt Differenzenoperatoren, welche aus einer endlichen Folge von Abbildungen konstruiert sind, kann man auch unendliche Folgen linearer Abbildungen betrachten. Der Raum $\mathrm{Hom}(V,W)$ der linearen Abbildungen von V nach W ist ebenfalls ein $\mathbb{C}$–Vektorraum. Auf diesem kann ein Skalarprodukt definiert werden. Seien $F, G \in \mathrm{Hom}(V,W)$. Die Abbildung $G^* \circ F : V \to V$ bildet nun einen endlichdimensionalen Vektorraum in sich ab. Somit kann von dieser Verknüpfung die Spur $\mathrm{spur}(G^* \circ F)$ gebildet werden. Wenn eine orthonormale Basis $\mathbf{e}_1, \ldots, \mathbf{e}_n \in V$ gewählt wird, hat die Spur eine einfache explizite Darstellung, es gilt

$$\mathrm{spur}(G^* \circ F) = \sum_{k=1}^{n} \langle\, (G^* \circ F)(\mathbf{e}_k),\, \mathbf{e}_k \,\rangle_V = \sum_{k=1}^{n} \langle\, F(\mathbf{e}_k),\, G(\mathbf{e}_k) \,\rangle_W \ .$$

Es ist einfach zu zeigen, dass die Spur von der Wahl der Orthonormalbasis unabhängig ist. Anhand des letzten Ausdrucks erkennt man, dass $\mathrm{spur}(F^* \circ F)$ immer nichtnegativ ist. Ist $\mathrm{spur}(F^* \circ F) = 0$, so muss das Bild jeden Basisvektors verschwinden, somit ist auch F die Nullabbildung.

Definition A.3.1 *Seien V ein endlichdimensionaler und W ein beliebiger $\mathbb{C}$–Hilbert–Raum. Auf dem Vektorraum $\mathrm{Hom}(V,W)$ definiert*

$$\mathrm{Hom}(V,W)^2 \ni (F,G) \mapsto \langle\, F,\, G \,\rangle_{\mathrm{Hom}(V,W)} := \mathrm{spur}(G^* \circ F)$$

ein Skalarprodukt, $\|F\|_{\mathrm{Hom}(V,W)} := \sqrt{\langle\, F,\, F \,\rangle_{\mathrm{Hom}(V,W)}}$ ist die vom Skalarprodukt induzierte Norm.

Lemma A.3.2 *Für beliebige $v \in V$ und $F : V \to W$ gilt $\|F(v)\|_W \leq \|F\|_{\mathrm{Hom}(V,W)}\, \|v\|_V$.*

Beweis: Sei $\mathbf{e}_1, \ldots, \mathbf{e}_n \in V$ eine Orthonormalbasis. Dann gilt

$$F(v) = \langle\, v,\, \mathbf{e}_1 \,\rangle_V\, F(e_1) + \cdots + \langle\, v,\, \mathbf{e}_n \,\rangle_V\, F(e_n) \ .$$

Mit der Cauchy–Schwarzschen Ungleichung folgt

$$\|F(v)\| \leq \sum_{k=1}^{n} |\, \langle\, v,\, \mathbf{e}_k \,\rangle_V \,|\ \|F(e_k)\|_W \leq \|v\|_V \|F\|_{\mathrm{Hom}(V,W)} \ .$$

□

Satz A.3.3 *Seien V ein endlichdimensionaler und W ein beliebiger $\mathbb{C}$–Hilbert–Raum. Ist $f \in \ell_1(\mathrm{Hom}(V,W))$ eine normsummierbare Folge linearer Abbildungen, so ist der* Differenzenoperator mit unendlichem Träger *$F := \sum_{n\in\mathbb{Z}} f_n \mathcal{T}^n$ für jedes $p \in [1,\infty)$ ein beschränkter linearer Operator der Folgenräume $F : \ell_p(V) \to \ell_p(W)$.*

Beweis: Wie für Differenzenoperatoren mit beschränktem Träger ist F nach Konstruktion linear und $(1,1)$–periodisch. Nun gilt für jede einzelne Abbildung $f_k : V \to W$, wenn sie auf Folgen $a \in \ell_p(V)$ ausgedehnt wird, weiterhin

$$\|f_k(a)\|_{\ell_p} \leq \sqrt[p]{\sum_{n\in\mathbb{Z}} \|f_k\|^p_{\mathrm{Hom}(V,W)} \|a_n\|^p_V} = \|f_k\|_{\mathrm{Hom}(V,W)} \|a\|_{\ell_p} \ .$$

Für den Differenzenoperator bedeutet dies, dass er beschränkt ist, denn es folgt

$$\|F(a)\|_{\ell_p} \leq \sum_n \|f_n \mathcal{T}^n(a)\|_{\ell_p} \leq \|f\|_{\ell_1} \|a\|_{\ell_p} .$$

□

Anhang B

Einige Grundbegriffe der Fourier–Analysis

Die trigonometrische Fourier–Reihe zu einer beliebigen Folge $c = \{c_n\}_{n\in\mathbb{Z}} \subset \mathbb{C}$ ist definiert als

$$\mathcal{E}[c] := \sum_{k\in\mathbb{Z}} c_n e_n, \tag{B.1}$$

wobei $e_n : \mathbb{R} \to \mathbb{C}$ für ein beliebiges $n \in \mathbb{Z}$ die Funktion $x \mapsto e_n(x) := e^{i\,2\pi n\,x}$ bezeichnet. Diese Reihe konvergiert gegen eine stetige Funktion, wenn die Folge c endlich ist, d.h. fast alle ihrer Glieder Null sind, oder, wie eben diskutiert, wenn die Reihe der Beträge der Glieder von c konvergiert, d.h. $c \in \ell_1(\mathbb{Z},\mathbb{C})$ gilt. Es soll im Folgenden die allgemeine Konvergenz dieser Reihe in den Rahmen der Theorie der Orthonormalsysteme in Hilbert–Räumen gestellt werden. Dabei ergibt sich, dass das Funktionensystem $\{e_n\}_{n\in\mathbb{Z}}$ sogar eine Hilbert–Basis im Funktionenraum $L^2([-\frac{1}{2},\frac{1}{2}])$ der über dem Intervall $[-\frac{1}{2},\frac{1}{2}]$ quadratintegrierbaren Funktionen ist.

B.1 Das Orthonormalsystem der trigonometrischen Monome

Sei mit I das Intervall $[-\frac{1}{2},\frac{1}{2}]$ bezeichnet. Wir betrachten die Funktionenfamilie $\{e_k\}_{k\in\mathbb{Z}}$, $e_k(x) := e^{i2\pi kx}$, als Teilmenge des komplexen Hilbert–Raumes $L^2(I) := L^2(I,\mathbb{C})$ mit dem üblichen Skalarprodukt

$$\langle\, f,\, g\,\rangle := \int_I f(x)\overline{g(x)}\,\mathrm{d}x\,, \quad \forall f,g \in L^2(I)\,.$$

Bezüglich dieses Skalarprodukts bildet die Familie $\{e_k\}_{k\in\mathbb{Z}}$ ein Orthonormalsystem, denn es gilt

$$\langle\, e_k,\, e_m\,\rangle = \int_I e^{i\,2\pi k\,x}\, e^{-i\,2\pi m\,x}\,\mathrm{d}x = \mathrm{sinc}(k-m) = \begin{cases} 1 & k = m \\ 0 & k \neq m \end{cases}.$$

Somit definiert der Fourier–Operator $\mathcal{E} : \ell_2(\mathbb{Z}) \to L^2(I)$ eine isometrische Einbettung des Hilbert–Raums $\ell_2(\mathbb{Z}) := \ell_2(\mathbb{Z},\mathbb{C})$ der quadratsummierbaren Folgen in den Hilbert–Raum $L^2(I)$ der quadratintegrierbaren Funktionen mit Träger in I. Es gilt $\|\mathcal{E}(c)\|_2 = \|c\|_2$ für jedes $c \in \ell_2(\mathbb{Z})$. In umgekehrter Richtung weiß man, ohne weitere Betrachtungen, lediglich, dass der adjungierte Operator $\mathcal{E}^* : L^2(I) \to \ell_2(\mathbb{Z})$ die Form

$$f \mapsto \mathcal{E}^*(f) = \{\langle\, f,\, e_n\,\rangle_{L^2(I)}\}_{n\in\mathbb{Z}}$$

hat, und ebenfalls die Operatornorm 1 besitzt, d.h. $\|\mathcal{E}^*(f)\| \leq \|f\| \ \forall f \in L^2(I)$. Ausgeschrieben ist dies die *Bessel–Ungleichung* für Orthonormalsysteme:

$$\sum_{n\in\mathbb{Z}} |\langle f, e_n \rangle|^2 \leq \|f\|^2 \quad \forall f \in L^2(I) \, .$$

B.2 Approximation der Einheit

Wir wissen bereits nach Satz A.1.1, dass wir Folgen $a \in \ell_2(\mathbb{Z})$ durch Folgen der Form $b(a) := \{b_n a_n\}_{n\in\mathbb{Z}} \in \ell_1(\mathbb{Z})$ approximieren können, wobei $b \in \ell_1(\mathbb{Z}, [0,1])$. Modifizieren wir auf diese Art die Folge $\mathcal{E}^*(f)$ der Fourier–Koeffizienten einer Funktion $f \in L^2(I) \subset L^1(I)$, so gibt es zur Folge $b(\mathcal{E}^*(f)) \in \ell_1(\mathbb{Z})$ wieder eine absolut und gleichmäßig konvergente Fourier–Reihe. Aus demselben Grunde konvergiert $K := \sum_{n\in\mathbb{Z}} b_n e_n$ gegen eine stetige, 1–periodische Funktion. Wir können die „gedämpfte" Fourier–Reihe zu f nun folgendermaßen umformen:

$$\begin{aligned} \mathcal{E}(b(\mathcal{E}^*(f)))(x) &= \sum_{n\in\mathbb{Z}} b_n \, \langle f, e_n \rangle_{L^2(I)} \, e_n(x) \\ &= \int_I f(t) \sum_{n\in\mathbb{Z}} b_n \, e_n(x-t) \, \mathrm{d}t = (K * f)(t) \, . \end{aligned} \tag{B.2}$$

Dabei stimmt die Reihe der gliedweisen Integrale mit dem Integral über die Reihe wegen

$$\int_I \sum_{n\in\mathbb{Z}} |f(t) b_n \, e_n(x-t)| \, \mathrm{d}t \leq \|f\|_{L^1} \|b\|_{\ell_1} < \infty$$

überein.

Wir wissen aus Satz A.2.2, dass es Familien von Funktionen gibt, welche eine Approximation der Eins für die Faltung sind. Können wir eine solche Familie finden, deren Funktionen nach der Art der Funktion K gebildet sind, so kann damit die Identität einer Funktion in $L^2(I)$ mit ihrer Fourier–Reihe gezeigt werden.

Definition B.2.1 (vgl. Sätze A.1.1 und A.2.2, s. [BZ97, Off9X])
Es sei $b := \{b_n\}_{n\in\mathbb{Z}} \subset C([0,1],[0,1])$ eine Folge stetiger Funktionen mit $b_n = b_{-n}$ und $b_0 \equiv 1$.

Wir sagen, dass b eine Approximation der Einheit *erzeugt, wenn*

- *$b(\lambda) := \{b_n(\lambda)\} \in \ell_1(\mathbb{Z})$ für jedes $\lambda \in (0,1]$ gilt,*
- *$\lim_{\lambda\to 0} b_n(\lambda) = 1$ ist für jedes $n \in \mathbb{Z}$, d.h. $b(1) = \{1\}_{n\in\mathbb{Z}}$, und wenn*
- *die Funktionen $K_\lambda := \chi_I \mathcal{E}(b(\lambda))$, die für $x \in I$ durch*

$$K_\lambda(x) := \sum_{n\in\mathbb{Z}} b_n(\lambda) e^{i2\pi nx} = 1 + 2\sum_{n=1}^{\infty} b_n(\lambda) \cos(2\pi nx)$$

definiert sind, eine Approximation der Eins $\{K_\lambda : \ \lambda \in (0,1]\} \subset L^1(\mathbb{R})$ bilden.

Die Folge b wird dann als Summationsmethode *und die Funktionen K_λ als die zugehörigen* Faltungskerne *bezeichnet.*

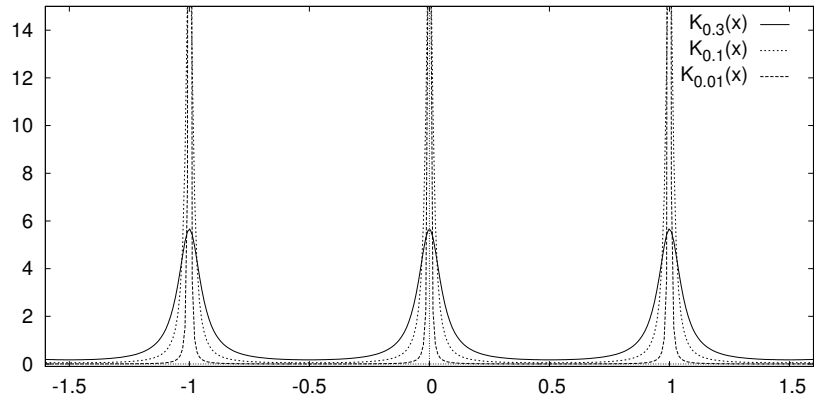


Abbildung B.1: Der Poisson–Kern K_λ für Werte $\lambda = 0.3, 0.1, 0.01$

Satz B.2.2 (Abel–Summation und Poisson–Kern)
Die Funktionenfolge $b = \{b_n\}_{n\in\mathbb{Z}}$ mit $b_n(\lambda) := (1-\lambda)^{|n|}$ erzeugt eine Approximation der Einheit. Mit $q := 1-\lambda$ gilt

$$K_\lambda(x) = \mathcal{E}(b(\lambda))(x) = \frac{1-q^2}{(1-q)^2 + 4q\sin^2 \pi x}\,.$$

Die Familie der Funktionen K_λ wird Poisson–Kern genannt.

Beweis: Die Funktionen b_n sind für jedes $n \in \mathbb{Z}$ stetig und bilden das Intervall $[0,1]$ auf sich ab, es gilt $b_n(0) = 1$ sowie $b_0 \equiv 1$. Nach Konstruktion ist b symmetrisch. Da die Folgenglieder nichtnegativ sind, gilt $\|b(\lambda)\|_{\ell_1} = \sum_{n\in\mathbb{Z}} b_n(\lambda) = K_\lambda(0)$. Zeigen wir also die Konvergenz von K_λ, so ist auch $b(\lambda) \in \ell_1(\mathbb{Z})$ gezeigt. Sei wie oben $q := 1-\lambda$ und weiterhin $z := qe^{i2\pi\, x}$ gesetzt. Es gilt $|z| < 1$ und damit

$$\begin{aligned} K_\lambda(x) =& 1 + \sum_{n=1}^{\infty} q^n (e^{i2\pi nx} + e^{-i2\pi nx}) = -1 + \sum_{n=0}^{\infty} (z^n + \bar{z}^n) = -1 + \frac{1}{1-z} + \frac{1}{1-\bar{z}} \\ =& \frac{1-|z|^2}{1+|z|^2 - 2Re(z)} = \frac{1-q^2}{(1-q)^2 + 2q(1-\cos(2\pi x))} = \frac{1-q^2}{(1-q)^2 + 4q\sin^2 \pi x} \end{aligned}$$

Damit ist $K_\lambda(0) = \frac{1+q}{1-q} = \frac{2-\lambda}{\lambda}$ endlich für $\lambda \in (0,1]$ und K_λ positiv und symmetrisch. Ist $\delta \in \left(0, \frac{1}{2}\right]$ gegeben, so ist $\sin^2 \pi x$ für $\delta \le |x| \le \frac{1}{2}$ durch $\sin^2 \pi\delta > 0$ nach unten beschränkt. Somit gilt

$$0 \le \int_{\delta\le|x|\le\frac{1}{2}} K_\lambda(x)\,\mathrm{d}x \le \frac{\lambda(2-\lambda)}{\lambda^2 + 4(1-\lambda)\sin^2 \pi\delta}\,.$$

Wird also δ konstant gehalten, so konvergiert das Integral bei $\lambda \to 0$ gegen Null. Da nach Konstruktion $\int_I K_\lambda(x)\,\mathrm{d}x = b_0(\lambda) = 1$ ist, bildet die Familie der $\varphi_\lambda := \chi_I K_\lambda$, $\lambda \in (0,1]$ eine Approximation der Eins. □

Analog dazu kann man zeigen, dass die ebenfalls klassische *Cesàro–Summation* eine Approximation der Einheit erzeugt.

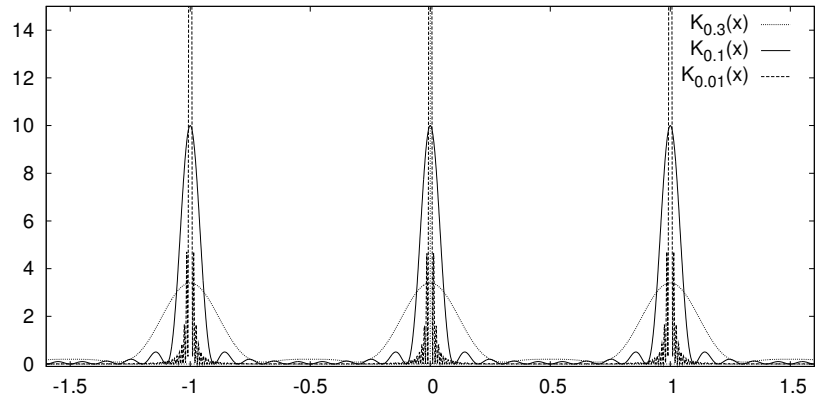


Abbildung B.2: Der Fejér–Kern K_λ für Werte $\lambda = 0.3, 0.1, 0.01$

Satz B.2.3 (Cesàro–Summation und Fejér–Kern)
Die Funktionenfolge $b = \{b_n\}_{n\in\mathbb{Z}}$ mit $b_n(\lambda) := (1-|n|\lambda)_+ := \max\big(0, 1-|n|\lambda\big)$ erzeugt eine Approximation der Eins. Mit $N \in \mathbb{N}$ derart, dass $\frac{1}{N+1} < \lambda \leq \frac{1}{N}$ gilt, ergibt sich

$$K_\lambda(x) = \mathcal{E}(b(\lambda))(x) = \frac{(1-\lambda N)\sin^2(\pi(N+1)x) + (\lambda(N+1)-1)\sin^2(\pi N x)}{\sin^2 \pi x}$$

Die Familie der Funktionen K_λ wird Fejér–Kern genannt.

Insbesondere ergibt sich für $\lambda = \frac{1}{N}$, dass $K_{\frac{1}{N}}(x) = \frac{\sin^2(\pi N x)}{N \sin^2 \pi x}$ gilt.

B.3 Konvergenz der trigonometrischen Fourier–Reihe

Ist $a := \mathcal{E}^*(f)$ die Folge der Fourier–Koeffizienten einer Funktion $f \in L^2(I)$, so wissen wir, dass mit der Bessel–Ungleichung des Orthonormalsystems $\{e_n : \ n \in \mathbb{Z}\}$ die Ungleichung $\|\mathcal{E}^*(f)\|_{\ell_2} \leq \|f\|_{L^2(I)}$ gilt. Mittels einer beliebigen Summationsmethode b kann nun auch die umgekehrte Ungleichung und damit die Gleichheit beider Seiten gezeigt werden.

Satz B.3.1 *Für jedes $f \in L^2(I)$, $I = \left[-\frac{1}{2}, \frac{1}{2}\right]$, konvergiert die Fourier–Reihe*

$$\mathcal{E}\mathcal{E}^*(f) = \sum_{n\in\mathbb{Z}} \langle\, f, e_n \,\rangle\, e_n$$

in der Norm von $L^2(I)$ gegen f. D.h. $\{e_n : \ n \in \mathbb{Z}\} \subset L^2(I)$ ist eine Hilbert–Basis.

Beweis: Seien ein $f \in L^2(I)$ und eine beliebige Summationsmethode b fixiert. Sei $a := \mathcal{E}^*(f)$ die Folge der Fourier–Koeffizienten bzgl. des Orthonormalsystems $\{e_n : \ n \in \mathbb{Z}\}$. Für jedes $\lambda \in (0,1]$ kann nun die „gedämpfte" Folge $a_\lambda := \{b_n(\lambda)\, a_n\}_{n\in\mathbb{Z}}$ konstruiert werden. Nach Satz A.1.1 gilt $a_\lambda \in \ell_1(\mathbb{Z})$ und $\|a_\lambda\|_{\ell_2} \leq \|a\|_{\ell_2}$, sowie die Konvergenz $\lim_{\lambda\to 0} \|a_\lambda - a\|_{\ell_2} = 0$.

Nach den Überlegungen zur „gedämpften" Fourier–Reihe in Gleichung (B.2) gilt $\mathcal{E}(a_\lambda) = K_\lambda * f$. Von dieser periodischen Funktion interessiert uns zur Bestimmung des Abstandes zu f in $L^2(I)$ nur derjenige Teil, welcher über dem Intervall $I = \left[-\frac{1}{2}, \frac{1}{2}\right]$ liegt. Mit $\varphi_\lambda := \chi_I K_\lambda$

und daher $K_\lambda = \sum_{n\in\mathbb{Z}} \mathcal{T}_n \chi_\lambda$ erhalten wir für $x \in I$

$$\begin{aligned}(K_\lambda * f)(x) &= \int_{\mathbb{R}} K_\lambda(t) f(x-t)\,\mathrm{d}t = \sum_{n\in\mathbb{Z}} \int_I \varphi_\lambda(n+t) f(x-n-t)\,\mathrm{d}t \\ &= \sum_{n=-1}^{1} \int_I \varphi_\lambda(t) f(x-n-t)\,\mathrm{d}t = \left(\varphi_\lambda * \sum_{n=-1}^{1} \mathcal{T}_n f\right)(x)\,.\end{aligned}$$

Dabei konnten wir die Summation auf $n \in \{-1,0,1\}$ einschränken, da sonst wegen $x,t \in I$ die Summe $(x+n-t)$ außerhalb $\operatorname{supp} f \subset I$ liegt. Sei $\tilde{f} := \mathcal{T}_{-1} f + f + \mathcal{T}_1 f$, es gelten also $\tilde{f} \in L^2(\mathbb{R})$ und $\chi_I \tilde{f} = f$.

Da nach Definition die Familie $\{\varphi_\lambda : \lambda \in (0,1]\}$ eine Approximation der Einheit ist, konvergiert nach Satz A.2.2 die Funktionenfamilie $\varphi_\lambda * \tilde{f}$ für $\lambda \to 0$ in $L^2(\mathbb{R})$ gegen $\tilde{f}$. Damit konvergiert die Einschränkung dieser Faltungsprodukte auf das Intervall I gegen f, unter anderem gilt $\lim_{\lambda\to 0} \|K_\lambda * f\|_{L^2(I)} = \|f\|_{L^2(I)}$.

Nun erhält der Fourier–Operator $\mathcal{E} : \ell_2(\mathbb{Z}) \to L^2(I)$ die Norm, insbesondere gilt

$$\|a_\lambda\|_{\ell_2} = \|\mathcal{E}(a_\lambda)\|_{L^2(I)} = \|K_\lambda * f\|_{L^2(I)}\,.$$

Setzen wir nun alle Ungleichungen zusammen, so erhalten wir

$$\|K_\lambda * f\|_{L^2(I)} \le \|\mathcal{E}^*(f)\|_{\ell_2} \le \|f\|_{L^2(I)}\,.$$

Bei $\lambda \to 0$ konvergiert nun die untere gegen die obere Schranke, somit muss schon von Anfang an $\|\mathcal{E}^*(f)\|_{\ell_2} = \|f\|_{L^2(I)}$ gegolten haben. Da dies für beliebiges $f \in L^2(I)$ gilt, ist das System $\{e_n : n \in \mathbb{Z}\}$ eine Hilbert–Basis. □

B.4 Die kontinuierliche Fourier–Transformation

Sei $g \in L^2(I)$. Es ist ohne weiteres möglich, in der Formel der trigonometrischen Fourier–Koeffizienten für den Index auch eine beliebige reelle Zahl einzusetzen, es sei dadurch eine Funktion $f : \mathbb{R} \to \mathbb{C}$ definiert,

$$x \mapsto f(x) := \langle\, g,\, e_x \,\rangle_{L^2(I)} = \int_I g(\omega) e^{-i(2\pi\omega)x}\,\mathrm{d}\omega\,.$$

Man überzeugt sich leicht, dass mit $\{e_n\}_{n\in\mathbb{Z}}$ auch die Funktionensysteme $\{e_{a+n}\}_{n\in\mathbb{Z}}$ für jedes $a \in \mathbb{R}$ Hilbert–Basen von $L^2(I)$ sind. Somit hat g auch die Fourier–Entwicklungen

$$g = \sum_{n\in\mathbb{Z}} f(n+a) e_{n+a}\,. \tag{B.3}$$

Lemma B.4.1 *Sei $g \in C_c^2(\mathbb{R},\mathbb{C})$ eine zweimal stetig differenzierbare Funktion mit Träger in $I = \left[-\frac{1}{2}, \frac{1}{2}\right]$. Dann ist die Funktion $f : \mathbb{R} \to \mathbb{C}$, $x \mapsto f(x) := \langle\, g,\, e_x \,\rangle_{L^2(I)}$ stetig und in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ enthalten. Es gelten die Identitäten*

$$\|f\|_{L^2(\mathbb{R})} = \|g\|_{L^2(\mathbb{R})} \quad \textit{und} \quad g(\omega) = \int_{\mathbb{R}} f(x) e_\omega(x)\,\mathrm{d}x \quad \forall \omega \in \mathbb{R}\,.$$

Beweis: Das Skalarprodukt $\langle g, h \rangle_{L^2(I)}$ hängt stetig vom zweiten Argument ab. Über dem Intervall $I = \left[-\frac{1}{2}, \frac{1}{2}\right]$ ist die Funktionenfamilie $\{e_x : x \in \mathbb{R}\} \subset L^2(I)$ stetig vom Parameter $x \in \mathbb{R}$ abhängig, somit ist f stetig.

Für die zweite Ableitung von g gilt nach zweifacher partieller Integration die Identität

$$\langle g'', e_x \rangle_{L^2(I)} = \int_I g''(\omega) e^{-i(2\pi\omega)x} \, \mathrm{d}\omega = (i2\pi x)^2 \int_I g(x) e^{-i(2\pi\omega)x} \, \mathrm{d}\omega = -(2\pi x)^2 f(x) \, .$$

Somit können die Funktionswerte von f abgeschätzt werden zu

$$(1 + |x|^2)|f(x)| \leq \int_I (|g(\omega)| + |g''(\omega)|) \, \mathrm{d}\omega \, .$$

Das Integral auf der rechten Seite ist eine von x unabhängige endliche Konstante $M > 0$, somit gilt $|f(x)| \leq M(1 + |x|^2)^{-1}$. Diese Majorante von f ist sowohl in der ersten als auch in der zweiten Potenz integrierbar, somit gilt $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

Aus der Majorante von f ergibt sich gleichfalls, dass die Reihen

$$\sum_{n \in \mathbb{Z}} |f(a+n)| \quad \text{und} \quad \sum_{n \in \mathbb{Z}} |f(a+n)|^2$$

für $a \in [0, 1]$ gleichmäßig konvergieren. Daraus ergibt sich mit dem Satz von Lebesgue über die dominierte Konvergenz zum einen für die Fourier–Entwicklungen (B.3)

$$\begin{aligned} g(\omega) &= \int_0^1 g(\omega) \, \mathrm{d}a = \int_0^1 \sum_{n \in \mathbb{Z}} f(n+a) e_{n+a}(\omega) \, \mathrm{d}a \\ &= \sum_{n \in \mathbb{Z}} \int_n^{n+1} f(a) e_\omega(a) \, \mathrm{d}a = \int_{\mathbb{R}} f(x) e^{i(2\pi\omega)x} \, \mathrm{d}x \end{aligned}$$

und zum zweiten für die Parseval–Gleichungen der Fourier–Koeffizienten in (B.3)

$$\begin{aligned} \|g\|_{L^2} &= \int_0^1 \|g\|_{L^2} \, \mathrm{d}a = \int_0^1 \sum_{n \in \mathbb{Z}} |f(n+a)|^2 \, \mathrm{d}a \\ &= \sum_{n \in \mathbb{Z}} \int_n^{n+1} |f(a)|^2 \, \mathrm{d}a = \int_{\mathbb{R}} |f(x)|^2 \, \mathrm{d}x = \|f\|_{L^2} \, . \end{aligned}$$

□

Durch Stauchen des Definitionsbereichs bzw. Vergrößern der Periode der benutzten Fourier–Reihen kann diese Aussage auf beliebige zweifach stetig differenzierbare Funktionen mit kompaktem Träger übertragen werden. Ist $g \in C_c^2(\mathbb{R})$ mit Träger in $[-N, N]$ für ein $N > 0$, so hat die Funktion $\tilde{g} := \mathcal{D}_{2N} g$ mit $x \mapsto \tilde{g}(x) := g(2Nx)$ ihren Träger im Intervall $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Die zwei transformierten Funktionen $f, \tilde{f} : \mathbb{R} \to \mathbb{C}$ mit

$$x \mapsto f(x) := \langle g, e_x \rangle_{L^2(\mathbb{R})} \quad \text{bzw.} \quad x \mapsto \tilde{f}(x) := \langle \tilde{g}, e_x \rangle_{L^2(\mathbb{R})} = \langle \tilde{g}, e_x \rangle_{L^2(I)}$$

stehen in der Beziehung

$$\tilde{f}(x) = \int_I \tilde{g}(\omega) \, e^{-i(2\pi\omega)x} \, \mathrm{d}\omega = \int_{\mathbb{R}} g(2N\omega) e^{-i(2\pi\, 2N\omega)\frac{x}{2N}} \, \mathrm{d}\omega = \tfrac{1}{2N} f\left(\tfrac{x}{2N}\right) \, .$$

In umgekehrter Richtung gilt dann

$$g(x) = \tilde{g}(\tfrac{\omega}{2N}) = \int_{\mathbb{R}} \tilde{f}(x) e^{i(2\pi\frac{\omega}{2N})x} \,\mathrm{d}x = \int_{\mathbb{R}} f(y) e^{i(2\pi\omega)y} \,\mathrm{d}y \,.$$

Ebenso überzeugt man sich, dass $\|f\|_{L^2(\mathbb{R})} = \|g\|_{L^2(\mathbb{R})}$ gilt.

Definition B.4.2 *Auf $C_c(\mathbb{R})$ sei die* Fourier–Transformation $\mathcal{F} : C_c(\mathbb{R}) \to C(\mathbb{R})$ *definiert als*

$$\mathcal{F}(f) := \hat{f} \text{ mit } \omega \mapsto \hat{f}(\omega) := \int_{\mathbb{R}} f(x) e^{-i(2\pi\omega)x} \,\mathrm{d}x$$

und die „adjungierte“ Fourier–Transformation $\mathcal{F}^* : C_c(\mathbb{R}) \to C(\mathbb{R})$ *als*

$$\mathcal{F}^*(\hat{f}) := f \text{ mit } x \mapsto f(x) := \int_{\mathbb{R}} \hat{f}(\omega) e^{i(2\pi\omega)x} \,\mathrm{d}\omega \,.$$

Bezeichnet $\bar{f}$ die Funktion mit komplex konjugierten Werten, so gilt $\mathcal{F}^*(\bar{f}) = \overline{\mathcal{F}(f)}$. Die beiden Abbildungen sind somit im wesentlichen identisch.

Die Fourier–Transformation ist $\mathbb{C}$–linear und in der L^2–Norm beschränkt. Da $C_c^2(\mathbb{R})$ eine dichte Teilmenge von $L^2(\mathbb{R})$ ist, kann diese lineare Abbildung zu einer beschränkten linearen Abbildung $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ auf eindeutige Weise fortgesetzt werden. Desgleichen gilt für $\mathcal{F}^*$. Mehr noch, in diesem Kontext ist $\mathcal{F}^*$ auch wirklich die adjungierte Abbildung zu $\mathcal{F}$, denn für stetige Funktionen $f, g \in C_c(\mathbb{R})$ gilt

$$\begin{aligned} \langle\, FT(f),\, g \,\rangle_{L^2(\mathbb{R})} &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x)\, e^{-i(2\pi\omega)x} \,\mathrm{d}x\, \overline{g(\omega)} \,\mathrm{d}\omega \\ &= \int_{\mathbb{R}} f(x) \overline{\int_R g(\omega)\, e^{i(2\pi\omega)x} \,\mathrm{d}\omega} \,\mathrm{d}x = \langle\, f,\, \mathcal{F}^*(g) \,\rangle_{L^2(\mathbb{R})} \,. \end{aligned}$$

Die Vertauschung der Integrationsreihenfolge ist nach dem Satz von Fubini möglich, da der Integrand stetig und absolut integrabel ist. Wegen der Stetigkeit des Skalarprodukts kann diese Identität auf beliebige $f, g \in L^2(\mathbb{R})$ ausgedehnt werden.

Aus denselben Dichtheitsgründen setzt sich auch die Identität $\mathcal{F}^*(\mathcal{F}(g)) = g$, die punktweise für alle zweifach stetigen Funktionen $g \in C_c^2(\mathbb{R})$ mit kompaktem Träger gilt, auf ganz $L^2(\mathbb{R})$ fort. Mit $\mathcal{F}^* \circ \mathcal{F} = id_{L^2(\mathbb{R})}$ gilt aber auch für jedes $f \in L^2(\mathbb{R})$

$$\bar{f} = \mathcal{F}^*(\mathcal{F}(\bar{f})) = \mathcal{F}^*\left(\overline{\mathcal{F}^*(f)}\right) = \overline{\mathcal{F}(\mathcal{F}^*(f))} \,.$$

Somit ist $\mathcal{F}^*$ auch die inverse Abbildung zu $\mathcal{F}$, die Fourier–Transformation ist also eine unitäre Abbildung auf $L^2(\mathbb{R})$.

Diese Eigenschaft setzt die oben auf $C_c^2(\mathbb{R})$ erhaltene Isometrie der Fourier–Transformation auf den gesamten Raum fort. Die Isometrieeigenschaft wird auch *Plancherel–Identität* genannt.

Satz B.4.3 (Plancherel–Identität) *Für beliebige $f \in L^2(\mathbb{R})$ gilt*

$$\|\mathcal{F}(f)\|_{L^2(\mathbb{R})} = \|f\|_{L^2(\mathbb{R})}$$

Beweis: Dies folgt aus den schon bekannten Eigenschaften, z.B. gilt mit $\mathcal{F}^* = \mathcal{F}^{-1}$

$$\|f\|^2_{L^2(\mathbb{R})} = \langle\, \mathcal{F}^*(\mathcal{F}(f)),\, f\,\rangle_{L^2(\mathbb{R})} = \langle\, \mathcal{F}(f),\, \mathcal{F}(f)\,\rangle_{L^2(\mathbb{R})} = \|\mathcal{F}(f)\|^2_{L^2(\mathbb{R})}$$

□

B.5 Translation, Modulation und Dilatation

Es gibt einige einfache, häufig benutzte Operationen auf $L^2(\mathbb{R})$, deren Wirkung unter der Fourier–Transformation ebenso einfach dargestellt werden kann.

Definition B.5.1 *Als* Modulation *einer Funktion $f \in L^2(\mathbb{R})$ mit Frequenz $\alpha \in \mathbb{R}$ wird das Produkt $e_\alpha f \in L^2(\mathbb{R})$ bezeichnet. Dabei ist die Funktion $e_\alpha : \mathbb{R} \to \mathbb{C}$ definiert als $x \mapsto e_\alpha(x) := e^{i(2\pi\alpha)x}$. Es gilt $e_\alpha(\beta) = e_\beta(\alpha)$.*

Die Dilatation *(Streckung oder Stauchung) einer Funktion $f \in L^2(\mathbb{R})$ um einen Faktor $s > 0$ ist die Funktion $\mathcal{D}_s f \in L^2(\mathbb{R})$ definiert durch $(\mathcal{D}_s f)(x) := f(sx)$.*

Die Translation *(oder Verschiebung) einer Funktion $f \in L^2(\mathbb{R})$ um eine Weglänge $t \in \mathbb{R}$ ist die Funktion $\mathcal{T}_t f \in L^2(\mathbb{R})$ definiert durch $(\mathcal{T}_t f)(x) := f(x - t)$.*

Sei eine Funktion $f \in C_c(\mathbb{R})$ fixiert und $\hat{f} = \mathcal{F}(f) \in \mathcal{L}^2(\mathbb{R}) \cap C(\mathbb{R})$ ihre Fourier–Transformierte.

Bei Modulation von f mit einer Frequenz $\alpha \in \mathbb{R}$ erhalten wir eine Translation von $\hat{f}$, $\mathcal{F}(e_\alpha f) = \mathcal{T}_\alpha \mathcal{F}(f)$. Denn für ein beliebiges $x \in \mathbb{R}$ gilt

$$\begin{aligned}(e_\alpha f)(x) &= \int_{\mathbb{R}} \hat{f}(\omega) e_{\omega+\alpha}(x)\, \mathrm{d}\omega = \int_{\mathbb{R}} \hat{f}(\omega - \alpha) e_\omega(x)\, \mathrm{d}\omega \\ &= \int_{\mathbb{R}} (\mathcal{T}_\alpha \hat{f})(\omega) e_\omega(x)\, \mathrm{d}\omega\,. \end{aligned} \tag{B.4a}$$

Für eine Translation von f um einen Abstand $t \in \mathbb{R}$ erhalten wir analog dazu eine Modulation von $\hat{f}$, $\mathcal{F}(\mathcal{T}_t f) = e_{-t}\mathcal{F}(f)$:

$$\begin{aligned}(\mathcal{T}_t f)(x) = f(x - t) &= \int_{\mathbb{R}} \hat{f}(\omega) e_\omega(x - t)\, \mathrm{d}\omega = \int_{\mathbb{R}} \hat{f}(\omega) e_\omega(-t) e_\omega(x)\, \mathrm{d}\omega \\ &= \int_{\mathbb{R}} (e_{-t}\hat{f})(\omega) e_\omega(x)\, \mathrm{d}\omega\,. \end{aligned} \tag{B.4b}$$

Eine Dilatation mit Faktor s wirkt sich als Dilatation mit reziprokem Faktor $\frac{1}{s}$ auf $\hat{f}$ aus, $\mathcal{F}(\mathcal{D}_s f) = \frac{1}{s}\mathcal{D}_{\frac{1}{s}}\mathcal{F}(f)$:

$$\begin{aligned}(\mathcal{D}_s f)(x) = f(sx) &= \int_{\mathbb{R}} \hat{f}(\omega) e_{s\omega}(x)\, \mathrm{d}\omega = \int_{\mathbb{R}} \hat{f}(\tfrac{\omega}{s}) e_\omega(x)\, \tfrac{1}{s}\, \mathrm{d}\omega \\ &= \int_{\mathbb{R}} \left(\tfrac{1}{s}\mathcal{D}_{\frac{1}{s}}\hat{f}\right)(\omega) e_\omega(x)\, \mathrm{d}\omega\,. \end{aligned} \tag{B.4c}$$

Anhang C

Systeme von Elementen eines Hilbert–Raumes

Die Rechteckfunktion $\chi_{[-\frac{1}{2},\frac{1}{2})}$ und der Kardinalsinus sinc stellen entgegengesetzte Extreme dar. Die eine Funktion ist Fourier–Transformierte der anderen. $\chi_{[-\frac{1}{2},\frac{1}{2})}$ besitzt einen kompakten Träger, ist aber nicht stetig. sinc ist analytisch, hat aber nur die sehr langsam im Unendlichen fallende einhüllende Kurve $\min(1, 1/|\pi x|)$. Beide Extreme sind für eine praktische Anwendung der Multiskalenanalyse ungeeignet. Wir suchen somit Funktionen, die sowohl selbst als auch in der Fourier–Transformierten schneller als $1/|x|$ fallen und mindestens stetig sind. Wir wollen nun untersuchen, unter welchen Bedingungen eine Funktion aus $L^2(\mathbb{R})$ eine Multiskalenanalyse des $L^2(\mathbb{R})$ erzeugt.

C.1 Motivation am endlichdimensionalen Hilbert–Raum

Wir bezeichnen mit System eine höchstens abzählbare Teilmenge eines Vektorraumes. Ist eine Funktion $\varphi \in L^2(\mathbb{R})$ gegeben, so können wir beispielsweise das System $\{\mathcal{T}^n\varphi : \ n \in \mathbb{Z}\}$ der Verschiebungen von φ betrachten. Endliche Linearkombinationen von Elementen dieses Systems sind immer in $L^2(\mathbb{R})$ enthalten. Eine erste Frage besteht darin, welche unendlichen Reihenentwicklungen mit solch einem System möglich sind. Weitere sind, ob aus dem Wert einer solchen Reihe auf die Koeffizienten, welche den einzelnen Elementen des Systems in ihr zugeordnet sind, geschlossen werden kann und ob diese Zuordnung eindeutig ist. Weiterhin kann man fragen, ob es zu jedem Element des Raumes eine Reihenentwicklung bzgl. dieses Systems gibt.

Ist in einem Hilbert–Raum ein Orthonormalsystem gegeben, so können wir zu diesem auf einfache Weise einen Projektor konstruieren, der jedem Element des Hilbert–Raumes die beste Approximation unter den Linearkombinationen von Vektoren des Orthonormalsystems zuordnet (s. Satz A.2.5).

Es ist jedoch nicht immer möglich oder wünschenswert, dass alle Vektoren einer abzählbaren Teilmenge des Hilbert–Raumes paarweise senkrecht zueinander stehen. Es wird im folgenden dargestellt, wie weit von der Orthogonalität abgewichen werden darf, so dass auf die oben gestellten Fragen trotzdem „sichere“ Antworten gegeben werden können.

Betrachten wir den endlichdimensionalen Fall. Im n–dimensionalen Spaltenvektorraum $V := \mathbb{K}^n$, $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$, mit dem kanonischen Skalarprodukt, definieren die Spalten einer

jeden $n \times m$–Matrix $\mathfrak{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_m)$ einen Unterraum von V, der von diesen aufgespannt wird.

Ist $m < n$ und hat die Matrix vollen Rang m, so definieren die Vektoren einen Unterraum. Nicht jeder Vektor kann also als Linearkombination dargestellt werden, aber es gibt zu jedem Vektor $\mathbf{v} \in V$ eine eindeutig bestimmte Linearkombination $\mathfrak{A}\mathbf{x}$, welche den minimalen Abstand

$$\|\mathbf{v} - \mathfrak{A}\mathbf{x}\|_2 = \min_{\mathbf{y}\in\mathbb{K}^m} \|\mathbf{v} - \mathfrak{A}\mathbf{y}\|_2$$

realisiert. Der Koeffizientenvektor ist $\mathbf{x} = (\mathfrak{A}^*\mathfrak{A})^{-1}\mathfrak{A}^*\mathbf{v}$, die Linearkombination $P(\mathbf{v}) := \mathfrak{A}\mathbf{x} = \mathfrak{A}(\mathfrak{A}^*\mathfrak{A})^{-1}\mathfrak{A}^*\mathbf{v}$ dazu entspricht der orthogonalen Projektion auf den von $\mathfrak{A}$ aufgespannten Unterraum. In der Tat gilt

$$P^2 = \mathfrak{A}(\mathfrak{A}^*\mathfrak{A})^{-1}\mathfrak{A}^*\mathfrak{A}(\mathfrak{A}^*\mathfrak{A})^{-1}\mathfrak{A}^* = P \quad \& \quad P^* = P\,.$$

Diese Situation entspricht dem Begriff eines *Riesz–Systems*, welcher insbesondere die Invertierbarkeit des auch im unendlichdimensionalen Fall definierten Operatorprodukts $\mathfrak{A}^*\mathfrak{A}$ zum Inhalt hat. Es kann also vom Wert einer Reihenentwicklung auf die Koeffizienten in der Reihe geschlossen werden.

Ist $m > n$ und hat $\mathfrak{A}$ vollen Rang n, so kann jeder Vektor in V durch Linearkombinationen aus Spalten von $\mathfrak{A}$ dargestellt werden, jedoch ist diese Darstellung nicht eindeutig. Man kann jedoch versuchen, einen gegebenen Vektor $\mathbf{v} \in V$ mit möglichst kleinem Koordinatenvektor darzustellen, d.h. unter allen $\mathbf{x} \in \mathbb{K}^m$ mit $\mathfrak{A}\mathbf{x} = \mathbf{v}$ denjenigen mit kleinstem Abstand $\|\mathbf{x}\|_2$ zum Ursprung zu finden. Nun ist bei gegebener Lösung $\mathbf{x}$ auch

$$P(\mathbf{x}) := \mathfrak{A}^*(\mathfrak{A}\mathfrak{A}^*)^{-1}\mathfrak{A}\mathbf{x}$$

eine Lösung, und da P ein orthogonaler Projektor ist, wird $P(\mathbf{x}) = \mathfrak{A}^*(\mathfrak{A}\mathfrak{A}^*)^{-1}\mathbf{v}$ sogar die kleinste Lösung. Diese Situation entspricht dem Begriff eines *Frames* oder *(aufspannenden) Vielbeins*. In dessen Definition wird insbesondere die Invertierbarkeit des Operatorprodukts $\mathfrak{A}\mathfrak{A}^*$ gesichert. Im endlichdimensionalen Fall bilden die Spalten der Matrix $(\mathfrak{A}\mathfrak{A}^*)^{-1}\mathfrak{A}$ den *dualen Frame*, die Skalarprodukte mit den Spaltenvektoren dieser Matrix ergeben gerade die „sparsamsten“ Koordinaten. Auch dies kann im unendlichdimensionalen Fall definiert werden und ergibt eine Reihenentwicklung eines jeden Vektors des Hilbert–Raumes bzgl. dieses Systems.

C.2 Bessel–Systeme

Bei einem unendlichdimensionalen Hilbert–Raum muss zunächst die Parametrisierbarkeit des von einer abzählbaren Teilmenge aufgespannten Unterraumes gesichert werden. Insbesondere muss geklärt werden, welche unendlichen Linearkombinationen im Hilbert–Raum konvergieren.

Seien $\mathcal{H}$ ein $\mathbb{K}$–Hilbert–Raum, $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$, im folgenden kurz Hilbert–Raum genannt, und $X \subset \mathcal{H}$ eine abzählbare Teilmenge darin. Mit $\ell_2(X) := \ell_2(X, \mathbb{K})$ sei die Menge aller durch X parametrisierten Folgen $c = \{c_x\}_{x\in X} \subset \mathbb{K}$ mit Werten im Skalarkörper von $\mathcal{H}$ bezeichnet,

für welche $\|c\|_{\ell_2}^2 := \sum_{x\in X} |c_x|^2 < \infty$ ist. Dieser Raum ist isometrisch isomorph zu $\ell_2(\mathbb{N}, \mathbb{K})$, dem Standard–Repräsentanten eines separablen Hilbert–Raums.

Definition C.2.1 *X heißt* Bessel–System, *falls eine* Bessel–Ungleichung *gilt, d.h. falls es eine Konstante $B > 0$ gibt mit*

$$\forall \mathbf{v} \in \mathcal{H} : \sum_{x\in X} |\langle \mathbf{v}, x \rangle_{\mathcal{H}}|^2 \leq B \|\mathbf{v}\|_{\mathcal{H}}^2 .$$

Wir können jedem Bessel–System X den linearen Operator

$$\mathcal{E}_X : \ell_2(X) \to \mathcal{H}, \quad c = \{c_x\}_{x\in X} \mapsto \mathcal{E}_X(x) := \sum_{x\in X} c_x \cdot x$$

und dessen adjungierten Operator

$$\mathcal{E}_X^* : \mathcal{H} \to \ell_2(X), \quad \mathbf{v} \mapsto \mathcal{E}_X^*(\mathbf{v}) := \{\langle \mathbf{v}, x \rangle_{\mathcal{H}}\}_{x\in X} ,$$

zuordnen. Beide Operatoren sind durch die Bessel–Konstante B beschränkt. Für $\mathcal{E}^*$ gilt dies nach Definition des Bessel–Systems und für die Operatornorm von $\mathcal{E}$ gilt nach Lemma A.2.4 $\|\mathcal{E}\| = \|\mathcal{E}^*\| < \infty$.

Nach [RS95] nennen wir $\mathcal{E}_X$ den *Synthese–* und $\mathcal{E}_X^*$ den *Analyse–Operator* des Systems X.

C.2.1 Verschiebungsinvariantes Bessel–System

Definition C.2.2 (vgl. [RS95]) *Sei $\Phi \subset L^2(\mathbb{R})$ eine höchstens abzählbare Teilmenge. Als von Φ erzeugtes* verschiebungsinvariantes System *bezeichnen wir die Menge*

$$X(\Phi) = \{\mathcal{T}^n \varphi : \varphi \in \Phi, n \in \mathbb{Z}\} \subset L^2(\mathbb{R}) .$$

Ist $\Phi = \{\varphi\}$ einelementig, so nennen wir dieses System $X(\varphi) := X(\Phi)$ von φ erzeugt.

Wir bezeichnen den Synthese–Operator von $X(\varphi)$ mit

$$\mathcal{E}_\varphi : \ell_{\text{fin}}(\mathbb{Z}) \to L^2(\mathbb{R}), \qquad c \mapsto \mathcal{E}_\varphi(c) := \sum_{n\in\mathbb{Z}} c_n \mathcal{T}_n \varphi$$

und den adjungierten Analyse–Operator mit

$$\mathcal{E}_\varphi^* : L^2(\mathbb{R}) \to \ell(\mathbb{Z}), \qquad f \mapsto \mathcal{E}_\varphi^*(f) := \{\langle f, \mathcal{T}^n \varphi \rangle\}_{n\in\mathbb{Z}} .$$

Betrachten wir ein $\varphi \in L^2(\mathbb{R})$ und das von φ erzeugte verschiebungsinvariante System $X(\varphi) = \{\mathcal{T}_n \varphi : n \in \mathbb{Z}\}$. Für jede endliche Folge $a \in \ell_{\text{fin}}(\mathbb{Z})$ gilt also $\mathcal{E}_\varphi(a) \in L^2(\mathbb{R})$, insbesondere können wir die Fourier–Transformierte dieser Funktionenreihe bilden, für diese gilt

$$\mathcal{F}(\mathcal{E}_\varphi(a)) = \sum_{n\in\mathbb{Z}} a_n \mathcal{F}(\mathcal{T}_n \varphi) = \left(\sum_{n\in\mathbb{Z}} a_n e_{-n} \right) \mathcal{F}(\varphi) .$$

Bezeichnen wir die Fourier–Reihe von a zum negativen Argument mit $\hat{a} := \sum_{n\in\mathbb{Z}} a_n e_{-n}$ und $\hat{\varphi} := \mathcal{F}(\varphi)$, so ist also $\mathcal{F}(\mathcal{E}_\varphi(a)) = \hat{a}\hat{\varphi}$. Die Fourier–Reihe $\hat{a}$ ist 1–periodisch. Mit den Rechteckfunktionen $\chi_m := \chi_{[m,m+1)}$ über den Intervallen $[m, m+1)$, $m \in \mathbb{Z}$, gilt

$$\hat{a}\hat{\varphi} = \sum_{m\in\mathbb{Z}} \chi_m \hat{a}\hat{\varphi} = \hat{a} \sum_{m\in\mathbb{Z}} \chi_m \hat{\varphi}\,.$$

Da die Intervalle disjunkt sind, sind die Summanden orthogonal. Die Norm von $\mathcal{E}_\varphi(a)$ kann somit bestimmt werden als

$$\begin{aligned}\left\|\mathcal{E}_\varphi(a)\right\|^2_{L^2(\mathbb{R})} &= \|\hat{a}\hat{\varphi}\|^2_{L^2(\mathbb{R})} = \sum_{m\in\mathbb{Z}} \|\hat{a}\chi_m\hat{\varphi}\|^2_{L^2(\mathbb{R})} \\ &= \int_0^1 |\hat{a}(\omega)|^2 \sum_{m\in\mathbb{Z}} |\hat{\varphi}(\omega+m)|^2 \,\mathrm{d}\omega\,. \end{aligned} \tag{C.1}$$

Die Funktion $G_\varphi : \mathbb{R} \to \mathbb{R}$, $G_\varphi(\omega) := \sum_{m\in\mathbb{Z}} |\hat{\varphi}(\omega+m)|^2$, ist messbar, $\int_I G_\varphi(\omega)\,\mathrm{d}\omega = \|\varphi\|^2 < \infty$. Nehmen wir an, dass $G_\varphi(\omega)$ (bis auf eine Nullmenge) durch ein $B > 0$ beschränkt ist. Dann erhalten wir eine Abschätzung

$$\left\|\mathcal{E}_\varphi(a)\right\|^2 \le B \int_0^1 |\hat{a}(\omega)|^2 \,\mathrm{d}\omega = B\|a\|^2_{\ell_2}\,. \tag{C.2}$$

Da diese Abschätzung von der Endlichkeit von a unabhängig ist, gilt sie auch für alle $a \in \ell_2(\mathbb{Z})$. Die Funktion $G_\varphi(\omega)$ hat die Form einer ℓ_2–Norm. Die darin vorkommenden Folgen von Werten der Fourier–Transformierten von φ sind ein wichtiges Werkzeug zur Analyse des von φ erzeugten verschiebungsinvarianten Systems.

C.2.2 Prä–Gramsche Fasern

Definition C.2.3 (vgl. [RS95, RS97b, Ron98]) *Sei $\Phi \subset L^2(\mathbb{R})$ eine höchstens abzählbare Teilmenge, jedes $\varphi \in \Phi$ hat eine mit $\hat{\varphi} := \mathcal{F}(\varphi) \in L^2(\mathbb{R})$ bezeichnete Fourier–Transformierte. Als* Prä–Gramsche Faser *von Φ wird die Funktion $J_\Phi : \mathbb{R} \to \mathrm{Hom}(\ell_{\mathrm{fin}}(\Phi), \ell_2(\mathbb{Z}))$ bezeichnet, welche für fast jedes $\omega \in \mathbb{R}$ eine lineare Abbildung $J_\Phi(\omega) : \ell_{\mathrm{fin}}(\Phi) \to \ell_2(\mathbb{Z})$ definiert,*

$$a = \{a_\varphi\}_{\varphi\in\Phi} \mapsto J_\Phi(\omega)(a) := \left\{ \sum_{\varphi\in\Phi} a_\varphi \hat{\varphi}(\omega+n) \right\}_{n\in\mathbb{Z}}\,.$$

Für eine einzelne Funktion $\varphi \in \ell_2$ ist also deren Prä–Gramsche Faser die Folge $J_\varphi(\omega) := \{\hat{\varphi}(\omega+n)\}_{n\in\mathbb{Z}}$.

Mittels der Prä–Gramschen Faser von Funktionen $f, g \in L^2(\mathbb{R})$ kann deren Skalarprodukt durch das $\ell_2(\mathbb{Z})$–Skalarprodukt ausgedrückt werden,

$$\langle f, g \rangle_{L^2(\mathbb{R})} = \int_0^1 \langle J_f(\omega), J_g(\omega) \rangle^2_{\ell_2} \,\mathrm{d}\omega\,. \tag{C.3}$$

Die wichtigste Eigenschaft der Prä–Gramschen Faser ist, dass sie die Trennung von Koeffizienten und erzeugender Funktion im Syntheseoperator eines verschiebungsinvarianten Systems

erlaubt. Sei $\Phi \subset L^2(\mathbb{R})$ höchstens abzählbar und $a \in \ell_{\text{fin}}(\Phi \times \mathbb{Z})$. Dann ist die Prä–Gramsche Faser zur Funktion $\mathcal{E}_\Phi(a)$ für jedes $\omega \in \mathbb{R}$ gegeben durch

$$J_{\mathcal{E}_\Phi(a)}(\omega) = \{ \sum_{\varphi \in \Phi} \hat{a}_\varphi(\omega + n) \hat{\varphi}(\omega + n) \}_{n \in \mathbb{Z}} = \sum_{\varphi \in \Phi} J_\varphi(\omega) \hat{a}_\varphi(\omega) = J_\Phi(\omega)(a(\omega)) \,. \tag{C.4}$$

Satz C.2.4 (vgl. [RS95] Satz 1.4.11) *Sei $\Phi \subset L^2(\mathbb{R})$ eine höchstens abzählbare Teilmenge. Das von dieser erzeugte verschiebungsinvariante System $X(\Phi) = \{\mathcal{T}_n \varphi : \varphi \in \Phi,\ n \in \mathbb{Z}\}$ ist genau dann ein Bessel–System mit einer Konstanten $B > 0$, wenn die Prä–Gramsche Faser von Φ (mit Ausnahme einer Menge vom Maß Null) für alle $\omega \in \mathbb{R}$ eine beschränkte lineare Abbildung*

$$J_\Phi(\omega) : \ell_2(\Phi) \to \ell_2(\mathbb{Z})$$

mit B als obere Schranke definiert.

Bemerkung: Ist $\Phi = \{\varphi\}$ einelementig, so bedeutet dies, dass die Prä–Gramsche Faser $J_\varphi(\omega)$ als Folge in $\ell_2(\mathbb{Z})$ für fast alle $\omega \in \mathbb{R}$ in der Kugel $K(0, \sqrt{B}) \subset \ell_2(\mathbb{Z})$ liegt, d.h. für diese Folge gilt

$$G_\varphi(\omega) := \|J_\varphi(\omega)\|_{\ell_2}^2 = \sum_{n \in \mathbb{Z}} |\hat{\varphi}(n + \omega)|^2 \leq B \,.$$

Beweis: Angenommen, $B > 0$ ist für fast alle $\omega \in \mathbb{R}$ eine obere Schranke der linearen Abbildung $J_\Phi(\omega)$. Analog zu Gleichung (C.2) und mit Gleichung (C.4) gilt dann für alle $\mathbf{a} \in \ell_{\text{fin}}(\Phi \times \mathbb{Z})$

$$\|\mathcal{E}_\Phi(\mathbf{a})\|_{L^2(\mathbb{R})}^2 = \int_0^1 \|J_\Phi(\omega)(\hat{\mathbf{a}}(\omega))\|^2 \, \mathrm{d}\omega \leq B \int_0^1 \|\hat{\mathbf{a}}(\omega)\|_{\ell_2(\Phi)}^2 \, \mathrm{d}\omega = B \|\mathbf{a}\|_{\ell_2(\Phi \times \mathbb{Z})} \,.$$

Somit gilt diese Abschätzung auch für alle $\mathbf{a} \in \ell_2(\Phi \times \mathbb{Z})$, und $X(\Phi)$ ist ein Bessel–System.

Sei nun angenommen, dass Φ ein verschiebungsinvariantes Bessel–System $X(\Phi)$ mit einer Konstanten B erzeugt. Für jedes Intervall $[c, d] \subset [0, 1]$ ist dessen charakteristische Funktion $\chi_{[c,d]} \in L^2([0,1])$ in eine Fourier–Reihe entwickelbar, d.h. es gibt eine Folge $\mathbf{a} = \{a_n\}_{n \in \mathbb{Z}} \in \ell_2(\mathbb{Z})$, deren Fourier–Reihe $\hat{\mathbf{a}} := \sum_{n \in \mathbb{Z}} a_n e_{-n}$ mit $\chi_{[c,d]}$ als Funktion in $L^2([0,1])$ übereinstimmt. Sei $\mathbf{b} \in \ell_2(\varphi)$ mit $\|\mathbf{b}\|_{\ell_2(\Phi)} \leq 1$. Dann ist deren Produktfolge $\mathbf{ab} := \{\mathbf{a}_n \mathbf{b}_\varphi\}_{(\varphi, n \in \Phi \times \mathbb{Z}}$ ein Element von $\ell_2(\Phi \times \mathbb{Z})$ mit Norm

$$\|\mathbf{ab}\|_{\ell_2(\Phi \times \mathbb{Z})}^2 = \sum_{(\varphi, n) \in \Phi \times \mathbb{Z}} |\mathbf{a}_n \mathbf{b}_\varphi|^2 \leq \sum_{n \in \mathbb{Z}} |\mathbf{a}_n|^2 = \|\hat{\mathbf{a}}\|_{L^2([0,1]}^2 = (d - c) \,.$$

Somit ist die Reihe $\mathcal{E}_\Phi(\mathbf{ab})$ in $L^2(\mathbb{R})$ enthalten und nach der Bessel–Ungleichung hat diese eine Normabschätzung

$$\|\mathcal{E}_\Phi(\mathbf{ab})\|_{L^2}^2 \leq B(d - c) \,.$$

Andererseits ist die Norm durch die Prä–Gramsche Faser ausdrückbar, es gilt $\widehat{\mathbf{ab}}(\omega) = \hat{\mathbf{a}}(\omega)\mathbf{b} = \chi_{[c,d]}(\omega)\mathbf{b} \in \ell_2(\Phi)$ und damit

$$\|\mathcal{E}_\Phi(\mathbf{ab})\|_{L^2}^2 = \int_0^1 \left\| J_\Phi(\omega) \left(\widehat{\mathbf{ab}}(\omega) \right) \right\|_{\ell_2(\Phi)}^2 \, \mathrm{d}\omega = \int_c^d \|J_\Phi(\omega)(\mathbf{b})\|_{\ell_2(\Phi)}^2 \, \mathrm{d}\omega \leq B(d - c) \,.$$

Da das Intervall $[c, d]$ und $\mathbf{b} \in \ell_2(\Phi)$ beliebig waren, muss $J_\Phi(\omega)$ durch B beschränkt sein. □

C.2.3 Gramsche und duale Gramsche Fasern

Sei $\varphi \in L^2(\mathbb{R})$ und das verschiebungsinvariante System $X(\varphi)$ ein Bessel–System. Mittels des Analyse–Operators $\mathcal{E}_\varphi^* : L^2(\mathbb{R}) \to \ell_2(\mathbb{Z})$ erhalten wir zu jeder Funktion $f \in L^2(\mathbb{R})$ eine quadratsummierbare Folge $\mathcal{E}_\varphi^*(f)$ von Skalarprodukten. Zu dieser Folge können wir die trigonometrische Fourier–Reihe, also $\widehat{\mathcal{E}_\varphi^*(f)} \in L^2([0,1])$, konstruieren. Diese können wir ebenso durch die Prä–Gramschen Fasern von φ und f ausdrücken. Nach Definition des adjungierten Operators muss für jede Folge $a \in \ell_2(\mathbb{Z})$ gelten

$$\begin{aligned}\left\langle \widehat{\mathcal{E}_\varphi^*(f)},\, \hat{a} \right\rangle_{L^2([0,1])} &= \left\langle \mathcal{E}_\varphi^*(f),\, a \right\rangle_{\ell_2} = \left\langle f,\, \mathcal{E}_\varphi(a) \right\rangle_{L^2(\mathbb{R})} = \int_0^1 \left\langle J_f(\omega),\, \hat{a}(\omega) J_\varphi(\omega) \right\rangle_{\ell_2} \mathrm{d}\omega \\ &= \left\langle \left\langle J_f,\, J_\varphi \right\rangle_{\ell_2},\, \hat{a} \right\rangle_{L^2([0,1])} .\end{aligned}$$

Damit ist $\widehat{\mathcal{E}_\varphi^*(f)} = \left\langle J_f,\, J_\varphi \right\rangle_{\ell_2}$, d.h. für fast jedes $\omega \in \mathbb{R}$ gilt

$$\widehat{\mathcal{E}_\varphi^*(f)}(\omega) = \left\langle J_f(\omega),\, J_\varphi(\omega) \right\rangle_{\ell_2} = \sum_{n\in\mathbb{Z}} \hat{f}(\omega+n)\, \overline{\hat{\varphi}(\omega+n)} \ . \tag{C.5}$$

Mittels der Prä–Gramschen Faser eines verschiebungsinvarianten Bessel–Systems können zwei Familien selbstadjungierter Operatoren gebildet werden.

Definition C.2.5 (vgl. [RS95, RS97b, Ron98]) *Sei $\Phi \subset L^2(\mathbb{R})$ eine höchstens abzählbare Teilmenge, so dass $X(\Phi)$ ein verschiebungsinvariantes Bessel–System ist.*

Als Gramsche Faser *zu $X(\Phi)$ wird die Funktion $G_\Phi(\omega) : \mathbb{R} \to \mathrm{Hom}(\ell_2(\Phi), \ell_2(\Phi))$ bezeichnet, deren Werte die fast überall definierten selbstadjungierten Abbildungen $G_\Phi(\omega) := J_\Phi(\omega)^* J_\Phi(\omega)$ sind. D.h. jede Folge $\mathbf{b} = \{\mathbf{b}_\varphi\}_{\varphi\in\Phi} \in \ell_2(\Phi)$ wird für fast alle $\omega \in \mathbb{R}$ auf die Folge*

$$G_\Phi(\omega)(\mathbf{b}) := \left\{ \sum_{\varphi\in\Phi} \mathbf{b}_\varphi \left\langle J_\varphi(\omega),\, J_{\tilde{\varphi}}(\omega) \right\rangle \right\}_{\tilde{\varphi}\in\Phi} \in \ell_2(\Phi)$$

abgebildet.

Als duale Gramsche Faser *zu $X(\Phi)$ wird die Abbildung $\tilde{G}_\Phi : \mathbb{R} \to \mathrm{Hom}(\ell_2(\mathbb{Z}), \ell_2(\mathbb{Z}))$ bezeichnet, deren Werte die fast überall definierten selbstadjungierten Abbildungen $\tilde{G}_\Phi(\omega) := J_\Phi(\omega) J_\Phi(\omega)^*$ sind. D.h. jede Folge $\mathbf{a} = \{\mathbf{a}_n\}_{n\in\mathbb{Z}} \in \ell_2(\mathbb{Z})$ wird für fast alle $\omega \in \mathbb{R}$ auf die Folge*

$$\tilde{G}_\Phi(\omega)(\mathbf{a}) := \sum_{\varphi\in\Phi} \left\langle a,\, J_\varphi(\omega) \right\rangle_{\ell_2(\mathbb{Z})} J_\varphi(\omega) \in \ell_2(\mathbb{Z})$$

abgebildet.

C.3 Riesz–Systeme

Definition C.3.1 *Seien $\mathcal{H}$ ein Hilbert–Raum und X ein Bessel–System darin. X wird* Riesz–System *genannt, falls es eine weitere positive Konstante $0 < A \le B$ gibt, so dass die* Riesz–Bedingung

$$\forall c \in \ell_2(X) : \ A\, \|c\|_{\ell_2}^2 \le \| \sum_{x\in X} c_x \cdot x \|^2$$

erfüllt ist.

Insgesamt gilt also in einem Riesz–System die zweiseitige Abschätzung

$$\forall c \in \ell_2(X): \ \sqrt{A}\|c\|_{\ell_2} \leq \|\mathcal{E}_X(c)\|_{\mathcal{H}} \leq \sqrt{B}\|c\|_{\ell_2} .$$

Gilt sogar $A = B$, so können die Elemente von X so skaliert werden, dass $A = B = 1$ und damit das Riesz–System ein Orthonormalsystem ist. Daher ist im Allgemeinen die normierte Differenz $\frac{B-A}{B+A}$ ein Maß für die Abweichung vom orthogonalen Fall.

C.3.1 Projektion auf den erzeugten Unterraum

Lemma C.3.2 *Seien $\mathcal{H}$ ein $\mathbb{C}$–Hilbert–Raum und $Q : \mathcal{H} \to \mathcal{H}$ ein selbstadjungierter Operator, welcher nach oben wie unten beschränkt sei, d.h. es gebe Konstanten $0 < A < B$, so dass für jeden Vektor $\mathbf{v} \in \mathcal{H}$ gilt*

$$A\|\mathbf{v}\|_{\mathcal{H}}^2 \leq \langle\, \mathbf{v},\, Q\mathbf{v} \,\rangle_{\mathcal{H}} \leq B\|\mathbf{v}\|_{\mathcal{H}}^2 .$$

Dann ist Q invertierbar mit einem beschränkten inversen Operator Q^{-1}, es gilt $\|Q^{-1}\| \leq \frac{1}{A}$.

Beweis: Sei $\gamma \in \mathbb{R}$ und sei mit I der identische Operator auf $\mathcal{H}$ bezeichnet. Wir betrachten die Differenz $I - \gamma Q$. Für diese gilt

$$(1 - \gamma B)\|\mathbf{v}\|_{\mathcal{H}}^2 \leq \langle\, \mathbf{v},\, (I - \gamma Q)\mathbf{v} \,\rangle_{\mathcal{H}} \leq (1 - \gamma A)\|\mathbf{v}\|_{\mathcal{H}}^2 ,$$

daher kann die Operatornorm von $I - \gamma Q$ abgeschätzt werden zu

$$\|I - \gamma Q\| = \sup_{\mathbf{v}\in\mathcal{H}:\|\mathbf{v}\|\leq 1} \left| \langle\, \mathbf{v},\, \mathbf{v} - \gamma Q\mathbf{v} \,\rangle_{\mathcal{H}} \right| \leq \max(1 - \gamma A,\, \gamma B - 1) .$$

Diese obere Schranke nimmt für $\gamma := \frac{2}{A+B}$ den minimalen Wert $\frac{B-A}{B+A} < 1$ an. Somit konvergiert die Neumann-Reihe $\sum_{k=0}^{\infty}(I - \gamma Q)^k$ und wir erhalten

$$Q^{-1} = \gamma(I - (I - \gamma Q))^{-1} = \gamma \left(I + \sum_{k=1}^{\infty} (I - \gamma Q)^k \right)$$

mit daraus folgender Abschätzung der Operatornorm des inversen Operators

$$\|Q^{-1}\| \leq \gamma \frac{1}{1 - \|I - \gamma Q\|} = \frac{2}{A+B} \frac{A+B}{2A} = \frac{1}{A} .$$

□

Ist $X \subset \mathcal{H}$ ein Riesz–System im Hilbert–Raum $\mathcal{H}$, so ist die Verknüpfung $Q := \mathcal{E}_X^* \circ \mathcal{E}_X : \ell_2(X) \to \ell_2(X)$ ein selbstadjungierter Operator auf dem Hilbert–Raum $\ell_2(X)$. Für jedes $c \in \ell_2(X)$ gilt

$$A\,\|c\|_{\ell_2}^2 \leq \langle\, \mathcal{E}_X(c),\, \mathcal{E}_X(c) \,\rangle_{\mathcal{H}} = \langle\, c,\, Qc \,\rangle \leq B\,\|c\|_{\ell_2}^2 .$$

Mit Lemma C.3.2 ist Q invertierbar und die Verknüpfung $P_X := \mathcal{E}_X \circ Q^{-1} \circ \mathcal{E}_X^* : \mathcal{H} \to \mathcal{H}$ ist ein orthogonaler Projektor auf $\mathcal{H}$. Dieser weist jedem Element $\mathbf{v} \in \mathcal{H}$ den nächstliegenden Vektor $P_X(\mathbf{v}) = \mathcal{E}_X(c)$ mit Koordinatenfolge $c := (\mathcal{E}^*\mathcal{E})^{-1}\mathcal{E}^*(\mathbf{v})$ im von X aufgespannten Unterraum zu.

Ein besonderer Fall ist gegeben, wenn dieser Projektor die Identität auf $\mathcal{H}$ ist, d.h. durch X der gesamte Hilbert–Raum aufgespannt wird.

Definition C.3.3 *Ein Riesz–System heißt* Riesz–Basis *oder* stabile Basis, *wenn $\mathcal{H} = im(\mathcal{E})$ gilt, d.h. der gesamte Hilbert–Raum wird durch Linearkombinationen von X mit Koeffizienten in $\ell_2(X)$ aufgespannt.*

Bemerkung: Ein Orthonormalsystem ist ein spezielles Riesz–System, eine Hilbert–Basis ist also auch eine Riesz–Basis.

C.3.2 Verschiebungsinvariante Bessel–Systeme

In dem wichtigen Spezialfall eines von einer Funktion $\varphi \in L^2(\mathbb{R})$ erzeugten verschiebungsinvarianten Bessel–Systems $\{\mathcal{T}_n\varphi : n \in \mathbb{Z}\}$ kann analog zu Satz C.2.4 die Riesz–Bedingung durch die Gramsche Faser ausgedrückt werden.

Satz C.3.4 *Sei eine Funktion $\varphi \in L^2(\mathbb{R})$ gegeben. Diese erzeugt genau dann ein Riesz–System $\{\mathcal{T}_n\varphi : n \in \mathbb{Z}\}$ mit Konstanten $0 < A < B$, wenn die Gramsche Faser von φ fast überall Werte aus dem Intervall $[A, B]$ annimmt. D.h. (mit Ausnahme einer Menge vom Maß Null) gilt für alle $\omega \in \mathbb{R}$*

$$A \leq G_\varphi(\omega) = \sum_{n\in\mathbb{Z}} |\hat{\varphi}(n+\omega)|^2 \leq B\,.$$

Beweis: Aus Satz C.2.4 folgt die Äquivalenz für die obere Schranke. Sei also vorausgesetzt, dass $\{\mathcal{T}_n\varphi : \ n \in \mathbb{Z}\}$ ein Bessel–System ist. Dann gilt für die Prä–Gramsche Faser einer Reihe zu einem $a \in \ell_2(\mathbb{Z})$ die Identität $J_{\mathcal{E}_\varphi(a)} = \hat{a}J_\varphi$. Gilt nun fast überall $A \leq G_\varphi(\omega) = \|J_\varphi(\omega)\|^2_{\ell_2}$, so erhalten wir eine Normabschätzung nach unten

$$\|\mathcal{E}_\varphi(a)\|^2_{L^2(\mathbb{R})} = \int_0^1 |\hat{a}(\omega)|^2\, \|J_\varphi(\omega)\|^2_{\ell_2}\, \mathrm{d}\omega \geq A\int_0^1 |\hat{a}(\omega)|^2\, \mathrm{d}\omega = A\|\hat{a}\|^2_{L^2([0,1])} = A\|a\|^2_{\ell_2}\,.$$

Erfüllt umgekehrt $\{\mathcal{T}_n\varphi : \ n \in \mathbb{Z}\}$ eine Riesz–Bedingung mit Konstanten $0 < A < B$, so gilt für jedes Intervall $[a, b] \subset [0, 1]$, dass $\chi_{[a,b]} \in L^2([0, 1])$ in eine Fourier-Reihe entwickelt werden kann, es gibt ein $a \in \ell_2(\mathbb{Z})$ mit $\hat{a} = \chi_{[a,b]}$, $\|a\|^2 = (b - a)$ und also

$$A(b-a) \leq \|\mathcal{E}_\varphi(a)\|^2_{L^2(\mathbb{R})} = \int_a^b \|J_\varphi(\omega)\|^2_{\ell_2}\, \mathrm{d}\omega \leq B(b-a)\,.$$

Somit muss außerhalb einer Menge vom Maß Null für alle $\omega \in \mathbb{R}$ gelten

$$A \leq \|J_\varphi(\omega)\|^2_{\ell_2} \leq B\,.$$

□

Für verschiebungsinvariante Systeme $X(\Phi)$, die von einem beliebigen System $\Phi \in L^2(\mathbb{R})$ erzeugt sind, kann dieser Satz ebenfalls formuliert werden. Um jedoch eine Verbindung zwischen Gramscher Faser und der Riesz–Bedingung zu knüpfen, muss die Gramsche Faser $G_\Phi(\omega)$ als selbstadjungierter Operator für fast jedes ω eine untere Schranke besitzen. Genauer, gibt es für diese Operatoren für fast alle $\omega \in \mathbb{R}$ gemeinsame obere und untere Schranken $0 < A \leq B$ mit

$$A\|a\|^2_{\ell_2(\Phi)} \leq \langle\, a,\, G_\Phi(\omega)(a)\,\rangle_{\ell_2(\Phi)} \leq B\|a\|^2_{\ell_2(\Phi)}\,,$$

so erzeugt Φ ein verschiebungsinvariantes Riesz–System und umgekehrt (s. [RS95], Satz 1.4.11).

C.4 Frames (Vielbein)

Definition C.4.1 *Seien $\mathcal{H}$ ein Hilbert–Raum und X ein Bessel–System mit einer Bessel–Konstanten $B > 0$ darin. X wird* Frame *genannt (engl. für „aufspannendes Vielbein"), falls es eine weitere positive Konstante A mit $0 < A \leq B$ gibt, so dass für jedes $\mathbf{v} \in \mathcal{H}$ gilt*

$$A \, \|\mathbf{v}\|_{\mathcal{H}}^2 \leq \sum_{x \in X} |\, \langle\, \mathbf{v},\, x \,\rangle\, |^2 \,. \tag{C.6}$$

Insgesamt gilt also in einem Frame die zweiseitige Abschätzung

$$\forall \mathbf{v} \in \mathcal{H} : \ \sqrt{A}\|\mathbf{v}\|_{\mathcal{H}} \leq \|\mathcal{E}_X^*(\mathbf{v})\|_{\ell_2} \leq \sqrt{B}\|\mathbf{v}\|_{\mathcal{H}} \,.$$

A und B werden *Frame–Schranken* genannt, gilt $A = B$, oder sogar $A = B = 1$, so wird der Frame straff (aufgespannt) (engl. „tight") genannt.

C.4.1 Dualer Frame

Ist X ein Frame in einem Hilbert–Raum $\mathcal{H}$, so ist der Operator $T := \mathcal{E}^*\mathcal{E} : \mathcal{H} \to \mathcal{H}$ selbstadjungiert. Nach Lemma C.3.2 gibt es für T einen beschränkten inversen Operator mit oberer Schranke $\frac{1}{A}$. Mittels des Analyse–Operators $\mathcal{E}_X^*$ kann jedem Vektor $\mathbf{v} \in \mathcal{H}$ eine Folge $c := \mathcal{E}_X^*(\mathbf{v}) \in \ell_2(X)$ zugewiesen werden. Die Invertierbarkeit von T ergibt nun, dass $(T^{-1}\mathcal{E}_X)(c) = \mathbf{v}$ gilt. Alternativ dazu kann eine Folge $\tilde{c} := (\mathcal{E}_X^* T^{-1})(\mathbf{v})$ gebildet werden, mit welcher man $\mathcal{E}_X(\tilde{c}) = \mathbf{v}$ erhält.

Definition C.4.2 *Sei $X \subset \mathcal{H}$ ein Frame, $\mathcal{E}_X : \mathcal{H} \to \ell_2(X)$ der Analyse–Operator und $R := (\mathcal{E}\mathcal{E}^*)^{-1}$. Dann nennt man $RX := \{Rx : \ x \in X\}$ den* dualen Frame *zu X.*

Nach Konstruktion kann ein vorgegebener Vektor $\mathbf{v} \in \mathcal{H}$ aus der Koeffizientenfolge $\tilde{c} := \mathcal{E}^* R(\mathbf{v})$ zurückgewonnen werden, denn es gilt $\mathcal{E}(\tilde{c}) = \mathcal{E}\mathcal{E}^*(\mathcal{E}\mathcal{E}^*)^{-1}(\mathbf{v}) = \mathbf{v}$. Diese Identität kann nach den einzelnen Elementen des Frames aufgelöst werden zu

$$\mathbf{v} = \sum_{x \in X} \langle\, \mathbf{v},\, Rx \,\rangle \cdot x \,. \tag{C.7}$$

Lemma C.4.3 *Das mit „dualer Frame" bezeichnete System ist ein Frame.*

Beweis: Für ein beliebiges $\mathbf{v} \in \mathcal{H}$ ist die Folge $(\mathcal{E}^* R)(\mathbf{v})$ sowohl nach oben als auch nach unten abzuschätzen. Es gilt zum einen

$$\|\mathcal{E}^*(R\mathbf{v})\|_{\ell_2}^2 = \langle\, (\mathcal{E}^*R)(\mathbf{v}),\, (\mathcal{E}^*R)(\mathbf{v}) \,\rangle_{\mathcal{H}} = \langle\, (\mathcal{E}\mathcal{E}^*R)(\mathbf{v}),\, R(\mathbf{v}) \,\rangle_{\mathcal{H}} = \langle\, \mathbf{v},\, R(\mathbf{v}) \,\rangle_{\mathcal{H}} \leq \frac{1}{A}\|\mathbf{v}\|_{\mathcal{H}}^2 \,.$$

Zum anderen verwenden wir, dass die Bessel–Ungleichung sowohl für $\mathcal{E}$ als auch für $\mathcal{E}^*$ gilt, woraus

$$\|\mathbf{v}\|_{\mathcal{H}}^2 = \|\mathcal{E}\mathcal{E}^* R\mathbf{v}\|_{\mathcal{H}}^2 \leq B\|\mathcal{E}^* R\mathbf{v}\|_{\ell_2}^2$$

folgt. □

Lemma C.4.4 *Sei $X \subset \mathcal{H}$ ein Frame in einem Hilbert–Raum. X ist eine* Riesz–Basis, *falls eine der folgenden äquivalenten Bedingungen gilt:*

i) $\mathcal{E}$ *ist injektiv,*

ii) $\mathcal{E}^*$ *ist surjektiv,*

iii) mit den Frame–Konstanten $0 < A \leq B < \infty$ *aus* (C.6) *gilt die Riesz–Bedingung*

$$\forall c \in \ell_2(X): \; A \, \|c\|_{\ell_2}^2 \leq \|\mathcal{E}(c)\|_{\mathcal{H}}^2 \leq B \, \|c\|_{\ell_2}^2 \, .$$

Beweis: Die behauptete Äquivalenz ist nachzuweisen. Eigenschaft iii) entspricht der Definition eines Riesz–Systems, da es ebenfalls ein Frame ist, ist es vollständig.

i) $\implies$ *ii)*: Für gegebenes $c \in \ell_2(X)$ ist die Gleichung $\mathcal{E}^*(f) = c$ zu lösen. Falls es eine Lösung $f \in \mathcal{H}$ gibt, so folgt $(\mathcal{E}\mathcal{E}^*)(f) = \mathcal{E}(c)$ und mit $R = (\mathcal{E}\mathcal{E}^*)^{-1}$ ergibt sich $f = (RT)(c)$.

Angenommen, die Gleichung $\mathcal{E}^*(f) = c$ ist für $f := (R\mathcal{E})(c)$ nicht erfüllt, dann ist die Differenz $d := c - \mathcal{E}^*(f) = c - (\mathcal{E}^* R \mathcal{E})(c) \neq 0 \in \ell_2(X)$, was $\mathcal{E}(d) = \mathcal{E}(c) - (\mathcal{E}\mathcal{E}^* R \mathcal{E})(c) = 0$ zur Folge hat, somit kann $\mathcal{E}$ nicht injektiv sein.

ii) $\implies$ *iii)* Die obere Abschätzung gilt, da ein Frame ein Bessel–System ist. Verbleibt die Abschätzung nach unten. Sei $c \in \ell_2(X)$ fest. Wegen der Surjektivität von $\mathcal{E}^*$ können wir $g \in \mathcal{H}$ so wählen, dass $\mathcal{E}^*(g) = \sqrt{A}c$ gilt. Nach der unteren Frame–Abschätzung (C.6) gilt

$$\sqrt{A}\|g\| \leq \|\mathcal{E}^*(g)\| = \sqrt{A}\|c\| \implies \|g\| \leq \|c\| \, .$$

und somit

$$\sqrt{A} \, \|c\|^2 = |\, \langle \, c, \, \mathcal{E}^* R \mathcal{E}(g) \, \rangle \, | = |\, \langle \, \mathcal{E}(c), \, g \, \rangle \, | \leq \|\mathcal{E}(c)\| \, \|g\| \leq \|\mathcal{E}(c)\| \, \|c\| \, .$$

iii) $\implies$ *i)* Auf Grund von $\|\mathcal{E}(c)\| \geq A \, \|c\|$ kann $\mathcal{E}(c) = 0$ nur für $c = 0$ erfüllt sein. □

C.4.2 Verschiebungsinvariante Frames

Man kann wieder untersuchen, ob ein von einem System $\Phi \subset L^2(\mathbb{R})$ erzeugtes verschiebungsinvariantes Bessel–System ein Frame ist. Zu diesem Zweck wird die duale Gramsche Faser $\tilde{G}_\Phi$ betrachtet. Ist für fast jedes $\omega \in \mathbb{R}$ der selbstadjungierte Operator $\tilde{G}_\Phi(\omega) : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ mit derselben Schranke A nach unten beschränkt, so kann ebenfalls gezeigt werden, dass das System $X(\Phi)$ die Frame–Bedingung erfüllt (s. [RS95]).

Bemerkung: Betrachten wir ein von einer Funktion $\varphi \in L^2(\mathbb{R})$ erzeugtes verschiebungsinvariantes System $\{\mathcal{T}_n \varphi\} \subset L^2(\mathbb{R})$, welches ein Frame für den von diesem System erzeugten Unterraum ist. Im Gegensatz zu einem Riesz–System kann dieser Unterraum im Allgemeinen nicht zur „Informationsübertragung“ genutzt werden, denn aus einer Linearkombination $\mathcal{E}_X(a) = \sum a_n \mathcal{T}_n \varphi$, $a \in \ell_2(\mathbb{Z})$, läßt sich die Folge a nicht eindeutig rekonstruieren. Jedoch kann

mit den Elementen des dualen Frames eine Funktion aus diesem Unterraum „gemessen“ werden und die gemessene Funktion aus der Messreihe rekonstruiert werden. Es gibt Beispiele, in welchen diese Rekonstruktion „robust“ gegen „kleine“ zufällige Störungen der Messreihe ist.

Literaturverzeichnis

[Abd96] ABDELJAOUED, J.: The Berkowitz Algorithm, Maple and Computing the Characteristic Polynomial in an Arbitrary Commutative Ring. (1996)

[Ber84] BERKOWITZ, S. J.: On Computing the determinant in small parallel time using a small number of processors. In: *Information Processing Letters* 18 (1984), S. 147–150

[BGHM01] BANK, B. ; GIUSTI, M. ; HEINTZ, J. ; MBAKOP, G.-M.: Polar Varieties and Efficient Real Elimination. In: *Mathematische Zeitschrift* 238 (2001), Nr. 1, S. 115–144

[BGHP05] BANK, Bernd ; GIUSTI, Marc ; HEINTZ, Joos ; PARDO, Luis M.: Generalized polar varieties: geometry and algorithms. In: *J. Complexity* 21 (2005), Nr. 4, S. 377–412

[BKN94] BETH, Thomas ; KLAPPENECKER, Andreas ; NÜCKEL, Armin: Construction of algebraic wavelet coefficients. In: *Proc. Int. Symp. on Information Theory and its Applications 1994*. Sydney, 1994 (ISITA'94), S. 341–344

[Bor99] BOREL, Emile: Mémoire sur ler séries divergentes. In: *Ann. Ecole Norm. Sup.* (1899)

[Buc65] BUCHBERGER, Bruno: *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. Österreich, Universität Innsbruck, Diss., 1965

[Buc79] BUCHBERGER, Bruno: A Criterion for Detecting Unnecessary Reductions in the Construction of Gröbner Bases. In: *Proc. EUROSAM 79* Bd. 72. Springer Verlag, 1979, S. 3–21

[Buc85] BUCHBERGER, B.: Gröbner Bases: An algoritmic method in polynomial ideal theory. In: BOSE ET AL, N. K. (Hrsg.): *Multidimensional System Theory*. Reidel, Dordrecht, 1985, S. 374–383

[BW93] BECKER, T. ; WEISPFENNING, B. V.: *Graduate Texts in Mathematics: readings in mathematics*. Bd. 141: *Göbner bases: a computational approach to commutative algebra*. New York : Springer Verlag, 1993

[BW99] BELOGAY, E. ; WANG, Y.: Construction of compactly supported symmetric scaling functions. In: *Applied and Computational Harmonic Analysis* 7 (1999), Nr. 2, 137–150. `http://citeseer.nj.nec.com/297084.html`

[BZ97] BENEDETTO, John J. ; ZIMMERMANN, Georg: Sampling multipliers and the Poisson summation formula. In: *Journal of Fourier Analysis and Applications* 3 (1997), Nr. 5, S. 505–523

[CLO97] COX, David ; LITTLE, John ; O'SHEA, Donal: *Ideals, varieties, and algorithms*. Second. New York : Springer-Verlag, 1997 (Undergraduate Texts in Mathematics). – xiv+536 S. – An introduction to computational algebraic geometry and commutative algebra

[CM99] CABRELLI, Carlos ; MOLTER, Ursula: Generalized Self-Similarity. In: *Journal of Mathematical Analysis and Applications* 230 (1999), S. 251–260

[Dau92] DAUBECHIES, Ingrid: *CBMS-NSF Regional Conference Series in Applied Mathematics*. Bd. 61: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics `http://www.siam.org/catalog/mcc02/daubechi.htm`

[Dau96] DAUBECHIES, Ingrid: Where do wavelets come from?—A personal point of view. In: *Proceedings of the IEEE Special Issue on Wavelets* 84 (1996), April, Nr. 4, S. 510–513

[Dem89] DEMAZURE, Michel: *Catastrophes et Bifurcations*. Paris : Ecole Polytechnique, 1989 (Editions Ellipses)

[Fau99] FAUGÈRE, Jean C.: A new efficent algorithm for computing Gröbner bases (F_4) / LIP6/CNRS Université Paris VI. 1999. – Forschungsbericht

[Ful84] FULTON, W.: *Intersection Theory*. 2. Springer, 1984 (Ergebnisse der Mathematik 3)

[GG86] GOLUBITSKY, M. ; GUILLEMIN, V.: *Graduate Texts in Mathematics*. Bd. 14: *Stable Mappings and their Singularities*. Third, corrected. Springer-Verlag, 1986

[GG99] GATHEN, Joachim von z. ; GERHARD, Jürgen: *Modern Computer Algebra*. Cambridge University Press, 1999

[GH80] GIUSTI, Marc ; HENRY, Jean–Pierre Georges: Minorations de nombres de Milnor. In: *Bulletin de la Société Mathématique de France* 108 (1980), S. 17–45

[GH91] GIUSTI, M. ; HEINTZ, J.: Algorithmes – disons rapides – pour la décomposition d' une variété algébrique en composantes irréductibles et équidimensionelles. In: MORA, T. (Hrsg.) ; TRAVERSO, C. (Hrsg.): *Proceedings of MEGA'90* Bd. 94, Birkhäuser, 1991, S. 169–194

[GHMP95] GIUSTI, M. ; HEINTZ, J. ; MORAIS, J. E. ; PARDO, L. M.: When Polynomial Equation Systems can be solved fast? In: COHEN, G. (Hrsg.) ; GIUSTI, H. (Hrsg.) ; MORA, T. (Hrsg.): *Applied Algebra, Algebraic Algorithms and Error Correcting Codes, Proceedings AAECC-11* Bd. 948, Springer, 1995, S. 205–231

[GLS01] GIUSTI, Marc ; LECERF, Grégoire ; SALVY, Bruno: A Gröbner Free Alternative for Polynomial System Solving. In: *Journal of Complexity* 17 (2001), Nr. 1, S. 154–211

[GM88] GEBAUER, R. ; MÖLLER, H. M.: On an Installation of Buchbergers Algorithm. In: *Journal of Symbolic Computation* 6 (1988), Nr. 2, 3, S. 275–286

[GRR99] GONZALES–VEGA, Laureano ; ROUILLIER, Fabrice ; ROY, Marie F.: Symbolic Recipes for Polynomial System Solving. In: COHEN, A. M. (Hrsg.) ; CUYPERS, H. (Hrsg.) ; STERK, H. (Hrsg.): *Some Tapas of Computer Algebra*, Springer–Verlag, 1999, S. 34–65

[Han98] HAN, Bin: Symmetric orthonormal scaling functions and wavelets with dilation factor 4. In: *Advances in Computational Mathematics* 8 (1998), Nr. 3, S. 221–247

[Hei79] HEINTZ, J.: Definability bounds of first order theories of algebraically closed fields. In: BUDACH, L. (Hrsg.): *Fundamentals of Computation Theory, FCT'97*. Berlin : Akademie–Verlag, 1979, S. 160–166

[Hei83] HEINTZ, Joos: Definability and fast quantifier elimination in algebraically closed fields. In: *Theoretical Computer Science* 24 (1983), S. 239–277

[Hel95] HELLER, Peter N.: Rank *M* Wavelets with *N* Vanishing Moments. In: *SIAM Journal on Matrix Analysis and Applications* 16 (1995), Nr. 2, S. 502–519

[Hig85] HIGGINS, John R.: Five short stories about the cardinal series. In: *Bull. Amer. Math. Soc.* 12 (1985), S. 45–89

[Hir64] HIRONAKA, Heisuke: Resolution of singularities of an algebraic variety over a field of characteristic zero. In: *Annals of Mathematics* 79 (1964), Nr. 1, S. 109–326

[Hir91] HIRSCH, Morris W.: *Graduate Texts in Mathematics*. Bd. 33: *Differential Topology*. Fourth, corrected. Springer-Verlag, 1991

[HMW01] HEINTZ, J. ; MATERA, G. ; WAISSBEIN, A.: On the time-space complexity of geometric elimination procedures. In: *Applicable Algebra in Engineering, Communication and Computing* 11 (2001), Nr. 4, S. 239–296

[HS80a] HEINTZ, J. ; SCHNORR, C. P.: Testing Polynomials which are easy to compute. In: *Proccedings of ACM 12th Symposium on Theory of Computing*, 1980, S. 262–272

[HS80b] HEINTZ, J. ; SIEVEKING, M.: Lower Bounds for Polynomials with algebraic coefficients. In: *Theo. Comp. Sci.* 11 (1980), S. 321–330

[HS81] HEINTZ, J. ; SIEVEKING, M.: Absolute primality of polynomials is decidable in random polynomial-time in the number of variables. In: *Proceedings ICALP 81* Bd. 115, Springer, 1981, S. 16–28

[Kal85] KALTOFEN, E.: Computing with polynomials given by straight-line programs I: Greatest common divisors. In: *Proceedings of the 17th Ann. ACM Symposium on Theory of Computing (Providence, RI)*. New York : ACM Press, 1985, S. 131–142

[Kro82] KRONECKER, L.: Grundzüge einer arithmetischen Theorie der algebraischen Grössen. In: *J. reine angew. Math.* 92 (1882), S. 1–122

[Kön03] KÖNIG, Julius: *Einleitung in die allgemeine Theorie der algebraischen Größen*. B. G. Teubner, Leipzig, 1903

[Lec00] LECERF, G.: Computing an Equidimensional Decomposition of an Algebraic Variety by means of Geometric Resolutions. In: *Proceedings of ISSAC'00* ACM, 2000

[Lec01a] LECERF, Grégoire: *Quadratic Newton Iteration for Systems with Multiplicity*. 2001

[Lec01b] LECERF, Grégoire: *Une alternative aux méthodes de réécriture pour la résolution des systèmes algébriques*, École polytechnique, Phd thesis, 2001. `http://tera.medicis.polytechnique.fr/tera/pub2001.html`. – Online–Ressource

[Lec03] LECERF, Grégoire: *Computing the Equidimensional Decomposition of an Algebraic Closed Set by means of Lifting Fibers*. 2003

[Leh99] LEHMANN, Lutz: *Effektives reelles Lösen einer multivariaten polynomialen Gleichung*, Humboldt–Universität zu Berlin, Diplomarbeit, 1999

[LW01] LEHMANN, L. ; WAISSBEIN, A.: Wavelets and semi–algebraic sets. In: *Proceedings of WAIT 2001* SADIO, 2001

[Mac16] MACAULEY, F. S.: *The Algebraic Theory of Modular Systems*. Cambridge University Press, 1916

[Mal89] MALLAT, Stephane G.: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-11 (1989), Nr. 7, S. 674–693

[Mat86] MATSUMURA, H.: *Commutative Ring Theory*. Cambridge University Press, 1986

[Mey89] MEYER, Yves: Orthonormal wavelets. In: COMBES, J. M. (Hrsg.) ; GROSSMAN, A. (Hrsg.) ; TCHAMITCHIAN, Ph. (Hrsg.): *Wavelets: Time-Frequency Methods and Phase Space*, Springer-Verlag, 1989, S. 21–37

[Mic91] MICCHELLI, Charles A.: Using the refinement equation for the construction of pre–wavelets. In: *Numerical algorithms* 1 (1991), S. 75–116

[MM82] MAYR, E. ; MEYER, A.: The complexity of the word problem for commutative semigroups. In: *Adv. in Math.* 46 (1982), S. 305–329

[Mor97] MORAIS, J. E.: *Resolución eficaz de sistemas de ecuaciones polinomiales*, Universidad de Cantabria, Santander, Spain, Diss., 1997

[Off9X] OFFNER, Carl D.: *A Little Harmonic Analysis*. `http://`. Version: 199X

[PS71] PÓLYA, G. ; SZEGÖ, G.: *Aufgaben und Lehrsätze aus der Analysis*. Bd. 2. Springer, 1971. – pp. 81–82 S. – VI, problems 40-43

[Ron98] RON, Amos: Wavelets and their associated operators. In: CHUI, C. K. (Hrsg.) ; SCHUMAKER, L. L. (Hrsg.): *Computational Aspects* Bd. 2, Vanderbilt University, 1998, S. 283–317

[Rou98] ROUILLIER, Fabrice: Solving zero–dimensional polynomial systems through the Rational Univariate Representation / INRIA Lorraine. 1998 (3426). – Forschungsbericht

[RS95] RON, Amos ; SHEN, Zuowei: Frames and stable bases for shift invariant subspaces of $L^2(R^d)$. In: *Canadian Journal of Mathematics* 47 (1995), Nr. 5, S. 1051–1094

[RS97a] RON, Amos ; SHEN, Zuowei: Affine systems in $L_2(\mathbb{R}^d)$ II: dual systems. In: *Journal of Fourier Analysis and Applications* 3 (1997), S. 617–637

[RS97b] RON, Amos ; SHEN, Zuowei: Affine systems in $L_2(\mathbb{R}^d)$: the analysis of the analysis operator. In: *Journal of Functional Analysis* 148 (1997), S. 408–447

[Sch80] SCHWARTZ, J. T.: Fast Probabilistic Algorithms for Verification of Polynomial Identities. In: *Journal of the ACM* 27 (1980), Oktober, Nr. 4, S. 701–717

[SF73] STRANG, G. ; FIX, G.: A Fourier analysis of the finite element variational method. In: GEYMONAT, G. (Hrsg.): *Constructive aspects of functional analysis*. Rome : Edizioni Cremonese, 1973

[Sha49] SHANNON, Claude E.: Communication in the presence of noise. In: *IEEE Journal on Signal Processing* (1949)

[SHGB93] STEFFEN, P. ; HELLER, P. ; GOPINATH, R. A. ; BURRUS, C. S.: Theory of regular *M*–band wavelet bases. 1993 (CML TR–93–02). – Forschungsbericht

[SZ98] STRANG, Gilbert ; ZHOU, Ding-Xuan: Inhomogeneous refinement equations. In: *Journal of Fourier Analysis and Applications* 4 (1998), S. 733–747

[Tur94] TURCAJOVÁ, Radka: Linear–Phase Paraunitary FIR Filter Banks: A Complete Characterization. In: LAINE, A. F. (Hrsg.) ; UNSER, M. (Hrsg.): *Mathematical Imaging: Wavelet Applications in Signal and Image Processing II, San Diego*, 1994 (SPIE Proceedings Vol. 2303), S. 47–55

[Vog84] VOGEL, W.: *Results on Bézout's Theorem*. Springer, 1984 (Tata Institute of Fundamental Research)

[vW31] VAN DER WAERDEN, B. L.: *Moderne Algebra, Zweiter Teil*. Die Grundlehren der modernen Wissenschaften in Einzeldarstellungen, 34. Julius Springer, Berlin, 1931

[Whi15] WHITTAKER, Edmund T.: On the functions which are represented by the expansion of interpolation theory. In: *Proc. R. Soc. Edinburgh* 35 (1915), S. 181–194

[Zip79] ZIPPEL, R.: Probabilistic algorithms for sparse polynomials. In: *Proceedings EUROSAM' 79*, Springer, 1979 (LNCS 72), S. 216–226

Index

Erklärung

Ich erkläre hiermit, diese Arbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben.

Es wird von mir derzeit keine weiterer Promotionsversuch unternommen.

Die Promotionsordnung der Mathematisch–Naturwissenschaftlichen Fakultät II ist mir bekannt.

Berlin, den